Biotechnology

Textbook for Class XI

© NCERT not to be republished

11150

विद्यया ऽमृतमश्नुते

एन सी ई आर टी
NCERT

राष्ट्रीय शैक्षिक अनुसंधान और प्रशिक्षण परिषद्
NATIONAL COUNCIL OF EDUCATIONAL RESEARCH AND TRAINING

Reprint 2026-27

11150 – BIOTECHNOLOGY
Textbook for Class XI

ISBN 978-93-5292-188-1

First Edition
October 2019 Ashwina 1941

Reprinted
August 2021 Shravana 1943
March 2022 Phalguna 1943
February 2025 Phalguna 1946

PD 3T RSP

© National Council of Educational Research and Training, 2019

₹ 330.00

Printed on 80 GSM paper with NCERT watermark

Published at the Publication Division by the Secretary, National Council of Educational Research and Training, Sri Aurobindo Marg, New Delhi 110 016 and printed at Indian Printing Works, E-4, Jhandewalan Ext., Rani Jhansi Road, New Delhi - 110 055

ALL RIGHTS RESERVED

- No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior permission of the publisher.
- This book is sold subject to the condition that it shall not, by way of trade, be lent, re-sold, hired out or otherwise disposed off without the publisher's consent, in any form of binding or cover other than that in which it is published.
- The correct price of this publication is the price printed on this page. Any revised price indicated by a rubber stamp or by a sticker or by any other means is incorrect and should be unacceptable.

OFFICES OF THE PUBLICATION DIVISION, NCERT

NCERT Campus
Sri Aurobindo Marg
New Delhi 110 016 **Phone : 011-26562708**

108, 100 Feet Road
Hosdakere Halli Extension
Banashankari III Stage
Bengaluru 560 085 **Phone : 080-26725740**

Navjivan Trust Building
P.O.Navjivan
Ahmedabad 380 014 **Phone : 079-27541446**

CWC Campus
Opp. Dhankal Bus Stop
Panihati
Kolkata 700 114 **Phone : 033-25530454**

CWC Complex
Maligaon
Guwahati 781 021 **Phone : 0361-2674869**

Publication Team

Head, Publication Division	: *M.V. Srinivasan*
Chief Editor	: *Bijnan Sutar*
Chief Production Officer (In charge)	: *Jahan lal*
Chief Business Manager	: *Amitabh Kumar*
Editor	: *Bijnan Sutar*
Production Assistant	: *Om Prakash*

Cover and Layout

DTP Cell, DESM

© NCERT not to be republished

Foreword

Biotechnology is comparatively a newer discipline as compared to Biology, Chemistry or Microbiology. It has emerged as a new subject to be taught in schools and colleges in the last two–three decades. As the name indicates, Biotechnology fundamentally deals with the application of laws and principles that govern and control the processes and phenomenon in living organisms.

Considering the fact that Biotechnology has a potential to provide solutions to many of the diverse problems that our society is facing right from protection and conservation of environment to treatment of diseases; production of alcoholic beverages to many humulin pharmaceutical products (one such example is monoclonal antibodies used for the treatment and diagnosis of diabetes); development of drought and disease resistant crop varieties in agriculture to genetically modified crops; understanding genetic bases of many of the phenomena happening in organisms to deciphering the whole genome and such others. All these have created new vistas and wider opportunities with tremendous potential.

This emerging area has not only helped in providing solutions to many problems and answers to a number of queries related to fundamentals of the processes and phenomena of living organisms, but, it has also opened the gate of interdisciplinary collaborations in newer areas. Today's Biotechnology or even Biology for that matter cannot be completely understood without the understanding of Physics and Chemistry. Similarly, generations of enormous data and its interpretation has opened up opportunities in yet another area called Bioinformatics, which largely depends on computer based applications, softwares and algorithms. This has a potential of even providing tailor made diagnosis and treatment of diseases and prediction of a person's possible suffering of diseases in future.

However, considering the fact that the present course is an entry level one, this book mainly focuses on understanding the fundamental concepts. Focus has also been given on problem solving skills by providing opportunities for hands-on activities and experiments in laboratories on one hand, and working on bioinformatics databases on the other.

Last but not the least, as an organisation committed to systemic reforms and continuous improvement in the quality of its teaching-learning products, NCERT has always welcomed comments and suggestions which enables us to improve the quality of materials. Valuable comments and suggestions on the book will also help NCERT to improve the content of the textbook.

New Delhi
November 2018

Director
National Council of Educational
Research and Training

© NCERT not to be republished

© NCERT
not to be republished

Preface

Quite recently, in the last two-three decades specialised disciplines like Biotechnology, Computer Science, Information Practices, etc., have emerged as priority areas in school education and these have been introduced at the higher secondary stage. This stage is challenging because of the transition from general to discipline-based curriculum. The higher secondary stage is also a connecting link between school education and higher and technical education. Therefore, syllabus at this stage needs to have appropriate rigour and depth while remaining mindful of the comprehension level of the learners. Further, the textbook need not be heavily loaded with content.

Biotechnology, as the name suggests, is an applied discipline which has potential to impact various facets. On one hand it has provided solutions to many health and medicine related problems, while on the other it has provided opportunities to explore newer areas like genomics, transcriptomics, proteomics, etc. These areas have implications to improve the quality of life besides solving many problems on various fronts like treatment of diseases, environmental protection and conservation, and understanding the process of evolution of life on earth, etc.

As an applied area related to molecular biology and biotechnology, Bioinformatics has also become a popular discipline due to generation of enormous amount of data in the area of genome biology. Needless to mention that about 15 years back we have seen the publication of draft human genome as an outcome of globally collaborated project called, 'Human Genome Project'.

Students take up Biotechnology with an aim of pursuing a career in molecular biology, molecular medicine, genome biology and various production industries related to biotechnology and molecular biology. Therefore, the course content of the subject must address all areas in which the subject has an implication. At the same time, it is also to be considered that the course must also make a foundation for higher and technical education. An attempt has been made in this direction to ensure that there is a balance between appropriateness and prospective need.

The course for Class XI has been divided into five units with 12 chapters. Unit I provides an introduction to the subject — its background and application in various areas. Unit II, has four chapters with details to understanding of cells, its bio-molecules including enzyme and cellular processes. Three chapters of Unit III will be helpful in developing the understanding of fundamentals of genetics, genetic material, mechanisms and processes related to DNA and RNA, and certain abnormalities in human beings especially related to chromosomal and genetic mechanisms. Unit IV, has three chapters, on quantitative biology, bioinformatics and programming in biology with application. The last unit of the book acquaints learners in the understanding of various tools and techniques used in the area of Biotechnology. An attempt has been made that the book provides a lucid reading to students and teachers so that it can effectively transact the concepts mentioned.

© NCERT not to be republished

I take this opportunity to place on record appreciation for U. N. Dwivedi, *Professor* Department of Biochemistry and Pro-Vice Chancellor of University of Lucknow, for leading the activities in the book as well as for his guidance and motivation to the development team. Thanks are due to the authors and reviewers for their valuable contribution.

Comments and suggestions towards the improvement of this book are welcome.

Dinesh Kumar
Professor and Head
Department of Education in
Science and Mathematics

© NCERT not to be republished

Reprint 2026-27

Textbook Development Committee

Chairperson

U. N. Dwivedi, *Professor*, Department of Biochemistry, University of Lucknow, Lucknow

Members

Amit Dinda, *Professor*, Department of Pathology, All India Institute of Medical Sciences, Delhi

Animesh Kumar Mohapatra, *Professor*, Regional Institute of Education, Bhubaneswar

Binay Panda, *Director,* Ganit Labs Foundation, Bio-IT Centre, Institute of Bioinformatics and Applied Biotechnology, Bengaluru

Indrakant K. Singh, *Assistant Professor*, Department of Zoology, Deshbandhu College, University of Delhi, Delhi

Kusum Yadav, *Assistant Professor*, Department of Biochemistry, University of Lucknow, Lucknow

Manoj K. Sharma, *Assistant Professor*, School of Biotechnology, Jawaharlal Nehru University, Delhi

Pawan K. Dhar, *Professor*, School of Biotechnology, Jawaharlal Nehru University, Delhi

Pushp Lata Verma, *Associate Professor*, Department of Education in Science and Mathematics, NCERT

Sunita Farkya, *Professor*, Department of Education in Science and Mathematics, NCERT

Member Co-ordinator

Dinesh Kumar, *Professor and Head*, Department of Education in Science and Mathematics, NCERT

© NCERT not to be republished

Acknowledgements

National Council of Educational Research and Training (NCERT) gratefully acknowledges the contribution of the individuals and organisations involved in the development of the Biotechnology textbook for Class XI. The Council is grateful to G. B. N. Chainy, *Professor*, Department of Zoology and Biotechnology, Utkal University, Odisha; Rupesh Chaturvedi, *Professor*, School of Biotechnology, Jawaharlal Nehru University, Delhi; Poonam Sharma, *Assistant Professor*, Gargi College, University of Delhi, Delhi; C. V. Shimray, *Assistant Professor*, Department of Education in Science and Mathematics, NCERT and Veda Prakash Pandey, *DST-SERB Young Scientist*, Department of Plant Biotechnology, Central Institute of Medicinal and Aromatic Plants (CIMAP), Lucknow for their contribution in the review of the manuscripts.

The Council is also thankful to Indrakant K. Singh, *Assistant Professor*, Deshbandhu College, University of Delhi, Delhi for providing pictures of electrophoresis apparatus and vertical section of maize and wheat leaf from his Molecular Biology Research lab.

NCERT is highly thankful to Archana Thakur, *Deputy Director*, Central Board of Secondary Education, Delhi; Shakun Singh, *PGT*, Bhatnagar International School, Vasant Kunj, Delhi; Madhumati Bhaskara, *PGT*, G.D. Goenka Public School, Vasant Kunj, Delhi; Anjulika Joshi, *PGT*, Mount Carmel School, Anand Niketan, Delhi; Payal Priyadarshini, *PGT*, Kendriya Vidyalaya, Delhi Cantt-3, Delhi; Pratibha Sharma, *PGT,* Kendriya Vidyalaya, JNU, Delhi and Ambika Nagrath, *PGT*, Army Public School, Dhaula Kuan, Delhi for their valuable suggestions.

Valuable suggestions and comments given by Ravindra Kumar Parashar, *Professor*, Department of Education in Science and Mathematics, NCERT and Alka Mehrotra, *Professor*, Department of Education in Science and Mathematics, NCERT especially on thermodynamics and biomolecules chapters have helped in improving the content of the book.

The Council also acknowledges the academic contributions of Priyal Sharma, *Junior Project Fellow*, in finalising the manuscript. Contributions of Suman Prajapati, *Graphic Designer* and Preeti Dhiman, *DTP Operator* for typesetting are also acknowledged. Without their effort it would not have been possible to bring out the manuscript. Cooperation from Rajendra Singh, *Assistant Program Coordinator*, and his staff for their help in organising workshops and office logistics for the same is especially thanked.

The efforts of Soumma Chandra, *Assistant Editor* (Contractual), C. Thangminlal Doungel, *Editorial Assistant* (Contractual), Chanchal Chouhan, *Proof Reader* (Contractual), and Naresh Kumar, *DTP Operator* (Contractual), of the Publication Division, NCERT in bringing out the first edition of this book is also highly appreciated.

© NCERT not to be republished

Contents

Foreword *iii*
Preface *v*

Unit I: An Introduction to Biotechnology 1-22

Chapter 1: Introduction 3

1.1 Historical Perspectives 4
1.2 Applications of Modern Biotechnology 8
1.3 Biotechnology in India: Academic Prospects and Industrial Scenario 16

Unit II: Cell Organelles and Biomolecules 23-144

Chapter 2: Cellular Organelles 25

2.1 Plasma Membrane 26
2.2 Cell Wall 29
2.3 Endomembrane System 31
2.4 Mitochondria 36
2.5 Plastids 37
2.6 Ribosomes 38
2.7 Microbodies 39
2.8 Cytoskeleton 40
2.9 Cilia and Flagella 41
2.10 Centrosome and Centrioles 42
2.11 Nucleus 43
2.12 Chromosome 45

Chapter 3: Biomolecules 50

3.1 Carbohydrates 50
3.2 Fatty Acids and Lipids 59
3.3 Amino Acids 63
3.4 Protein Structure 67
3.5 Nucleic Acids 75

Chapter 4: Enzymes and Bioenergetics 85

4.1 Enzymes: Classification and Mode of Action 85
4.2 Brief Introduction to Bioenergetics 96

© NCERT not to be republished

Chapter 5: Cellular Processes **103**

5.1 Cell Signaling 103
5.2 Metabolic Pathways 104
5.3 Cell Cycle 126
5.4 Programmed Cell Death (Apoptosis) 135
5.5 Cell Differentiation 136
5.6 Cell Migration 139

Unit III: Genetic Principles and Molecular Processes 145-232

Chapter 6: Basic Principles of Inheritance **147**

6.1 Introduction to Inheritance 147
6.2 Linkage and Crossing Over 153
6.3 Recombination 156
6.4 Sex-linked Inheritance 158
6.5 Extrachromosomal Inheritance 159
6.6 Polyploidy 160
6.7 Reverse Genetics 161

Chapter 7: Basic Processes **166**

7.1 DNA as the Genetic Material 166
7.2 Prokaryotic and Eukaryotic Gene Organisation 171
7.3 DNA Replication 175
7.4 Gene Expression 184
7.5 Genetic Code 191
7.6 Translation 193
7.7 Gene Mutation 199
7.8 DNA Repair 205
7.9 Regulation of Gene Expression 208

Chapter 8: Genetic Disorder **217**

8.1 Chromosomal Abnormalities and Syndromes 217
8.2 Monogenic Disorders and Pedigree Mapping 222
8.3 Polygenic Disorders 227

© NCERT not to be republished

Reprint 2026-27

Unit IV: Quantitative Biology and Bioinformatics 233-278

Chapter 9: Introduction to Bioinformatics 235

9.1 The Utility of Basic Mathematical and Statistical Concepts to Understand Biological Systems and Processes 235
9.2 Introduction 239
9.3 Biological Databases 244
9.4 Genome Informatics 247
9.5 Role of Artificial Intelligence (AI) in Future 253

Chapter 10: Protein Informatics and Cheminformatics 256

10.1 Protein Informatics 256
10.2 Cheminformatics 261

Chapter 11: Programming and Systems Biology 270

11.1 Programming in Biology 270
11.2 Systems Biology 272

Unit V: Tools and Techniques: Basic Concepts 279-316

Chapter 12: Tools and Techniques 281

12.1 Microscopy 281
12.2 Centrifugation 286
12.3 Electrophoresis 287
12.4 Enzyme-linked Immunosorbent Assay (ELISA) 291
12.5 Chromatography 294
12.6 Spectroscopy 297
12.7 Mass Spectrometry 301
12.8 Fluorescence in Situ Hybridisation (FISH) 301
12.9 DNA Sequencing 303
12.10 DNA Microarray 308
12.11 Flow Cytometry 311

© NCERT not to be republished

Reprint 2026-27

THE CONSTITUTION OF INDIA

PREAMBLE

WE, THE PEOPLE OF INDIA, having solemnly resolved to constitute India into a [1]**[SOVEREIGN SOCIALIST SECULAR DEMOCRATIC REPUBLIC]** and to secure to all its citizens :

JUSTICE, social, economic and political;

LIBERTY of thought, expression, belief, faith and worship;

EQUALITY of status and of opportunity; and to promote among them all

FRATERNITY assuring the dignity of the individual and the [2][unity and integrity of the Nation];

IN OUR CONSTITUENT ASSEMBLY this twenty-sixth day of November, 1949 do **HEREBY ADOPT, ENACT AND GIVE TO OURSELVES THIS CONSTITUTION.**

---

1. Subs. by the Constitution (Forty-second Amendment) Act, 1976, Sec.2, for "Sovereign Democratic Republic" (w.e.f. 3.1.1977)
2. Subs. by the Constitution (Forty-second Amendment) Act, 1976, Sec.2, for "Unity of the Nation" (w.e.f. 3.1.1977)

Reprint 2026-27

Unit I

An Introduction to Biotechnology

Chapter 1:
Introduction

Knowledge of natural sciences has been applied to develop technologies since long for the welfare and comfort of human beings. It has also contributed to enhance the value of human lives. Research in the fields of physics and chemistry gave rise to engineering and technology industries. Among the many fields of science is a broad area of biology called Biotechnology, which has now expanded to diverse fields such as genetics, immunology, agriculture, genomics, etc. This unit provides a comprehensive description to develop the understanding of early history of biotechnology along with the recent developments in this field.

© NCERT not to be republished

Karl Ereky (1878-1952)

The term biotechnology was coined by Karl Ereky, a Hungarian scientist, in his book entitled *Biotechnologie der Fleish-, Fett-und Milcherzeugung im landwirtschaftlichen Grossbetriebe* (Biotechnology of Meat, Fat and Milk Production in an Agriculture Large-scale Farm) in 1917.

In his book, he described how technology could be used to transform plants and animals into useful products.

© NCERT not to be republished

11150CH01

Chapter 1

Introduction

1.1 *Historical Perspectives*

1.2 *Applications of Modern Biotechnology*

1.3 *Biotechnology in India: Academic Prospects and Industrial Scenario*

Biotechnology, the term, is a combination of two words 'bio' and 'technology', — 'bio' means biological systems or processes, and 'technology' refers to methods, systems, and devices used to make useful products from these biological systems. Thus, biotechnology refers to the different technologies that make use of living cells and/or biological molecules to generate useful products for the benefit of mankind.

Mankind has been practicing biotechnology since long. Right from the domestication of sheep and cattle in the Paleolithic age, conservation of plant stocks by the early Egyptian farmers (ancient germplasm conservation), to the classical examples of early fermentation technology in the form of making bread, cheese and wine. However, modern biotechnology is a multidisciplinary subject which involves knowledge sharing between different areas of science such as Cell and Molecular Biology, Microbiology, Genetics, Anatomy and Physiology, Biochemistry, Computer Science and Recombinant DNA technology (rDNA technology).

© NCERT not to be republished

This chapter will elaborate on the history of biotechnological practices and the development of the modern concepts; major applications of biotechnology in the field of medicine, agriculture, food and environment conservation as well as the current scenario of the Indian biotechnology sector.

1.1 HISTORICAL PERSPECTIVES

Ancient biotechnology had taken root as early as in the Paleolithic era, around 10,000 years ago, when early farmers began to cultivate crops such as wheat and barley. Civilisations prevalent in the Sahara region of Africa were successfully domesticating sheep, goat and cattle, and were familiar with the techniques of hunting and the potential uses of fire. People collected the seeds of wild plants for cultivation and domesticated some species of wild animals living around them, executing, what is now known as 'selective breeding'. However, the most classical example of biotechnology in the medieval times is the use of fermentation technology for production of bread, cheese, wine and beer.

Science and traditional knowledge have always gone hand-in-hand in order to produce beneficial results. Greater efforts are being made to document and utilise the Indian traditional knowledge of medicine and biotechnology. People of ancient India had immense knowledge of their environment and properties of plants and animals. The practice of making fermented food such as *dahi*, *idli*, *kinema* and beverages using local biological resources was common in medieval India. The relevance of traditional Indian knowledge in making curd (*dahi*) has been indicated in few of the patents found in the United States patent database.

Fermentation can be explained as a microbial process in which enzymatically-controlled conversion of organic compounds occurs. Fermentation was practiced for years without any actual knowledge of the processes involved. Fermented dough was discovered by accident when dough was not baked immediately and consequently it underwent fermentation by yeast such as *Saccharomyces winlocki*. Egypt and Mesopotamia exported bread to Greece and Rome. In efforts to improve the technique, Baker's Yeast was discovered by the Romans, which revolutionised the

© NCERT not to be republished

Box 1

Curd Making: A Traditional Biotechnological Technique

We all must have observed our mothers making curd for the entire family. It is a classic example of fermentation technology, which can be conducted right at home.

Fresh Milk boiled to 100°C and cooled to 37°C

+

Add a teaspoonful of one-day-old curd and leave undisturbed for 4–5 hours

Two phase separation of curd and whey

Observation: The raw material, milk, has been converted completely into a semisolid sour tasting product.

WHAT HAS ACTUALLY HAPPENED?

Lactobacillus bacteria from the curd react with the milk protein casein. Lactic acid formed as a by-product, denatures the globular proteins and coagulates to produce the solid curds and separates the watery whey protein layer

bread-making technology prevalent then. The Chinese were also using fermentation technology by 4000 B.C., for production of their traditional food items, such as soy sauces and fermented vegetables. Vinegar production was known to the Egyptians by 2000 B.C., by preserving crushed dates for a longer time. The art of preserving animal foods by drying, smoking, and pickling in the brine were popular in pre-historic East and Europe.

Beer making may have begun as early as between 6000 and 5000 B.C. using cereal grains such as sorghum, corn, rice, millet, and wheat. Brewing was considered as an art until the fourteenth century A.D. However, early brewers had no practical knowledge about the microbial basis of fermentation. Wine was probably made by accident, when grape juice was contaminated with yeast and other microbes. Between 1850s and 1860s, Louis Pasteur established that yeast and other microbes were responsible for fermentation.

Nineteenth century witnessed an increase in the production scale of fermentation based products such as glycerol, acetone, butanol, lactic acid, citric acid, etc. Industrial fermentation was established during World War I because of Germany's requirement for large amounts of glycerol for explosives. By 1940's, significant

© NCERT not to be republished

improvement was made to techniques involving sterility maintenance, aeration methods, product isolation and purification. World War II was the catalyst which led to the invention of the modern fermenter (vessels used for fermentation), also called bioreactor for mass production of antibiotic penicillin. Today, many chemicals such as antibiotics, amino acids, hormones, pigments and even enzymes are produced with extreme precision in controlled environments of industrial bioreactors.

The foundation of modern biotechnology were laid down with the advancements in science and technology during the eighteenth and nineteenth centuries. Thus with the advent of the first compound microscope, made by Dutch spectacle-maker Zacharias Janssen in 1590, which could magnify about 3×-9×, enabled humans to 'see' things that were not perceived by naked eye.

In 1665, Robert Hooke, a physicist, examined thinly sliced cork and drew rectangular components, which he called *cellulae* (Latin for 'small chambers'). In 1676, Antonie van Leeuwenhoek, a Dutch shopkeeper, saw living organisms in pond water and called them 'animalcules'. During the eighteenth century, the cell theory was developed by German biologists, Matthias Schleiden and Theodor Schwann, who determined that all plant and animal tissues were composed of cells. In 1858, Rudolf Virchow, a German pathologist, concluded that 'all cells arise from pre-existing cells' and that cell is the basic unit of life.

Between 1850 and 1880, Pasteur developed the process of pasteurisation. By 1860, he also concluded that spontaneous generation of organisms did not occur, proving that 'all cells arise from pre-existing cells'. In 1896 Eduard Buchner converted sugar to ethyl alcohol using yeast extracts, showing that biochemical transformations can occur without the use of cells. By 1920s and 1930s, the biochemical reactions of many important metabolic pathways were established.

Genetics and Principles of Heredity was developed by an Austrian monk named Gregor Mendel, beginning in 1857, when he cross-pollinated pea plants to examine traits such as petal color, seed color, and seed texture. In 1869, Johann Friedrich Miescher, a

© NCERT not to be republished

Swiss biochemist, isolated a substance that he called nuclein from the nuclei of white blood cells. The substance contained nucleic acids. In 1882, German cytologist Walter Flemming described thread-like bodies that were visible during cell division, as well as the equal distribution of this material to daughter cells. These thread like bodies were actually chromosomes dividing between the two daughter cells during mitosis.

Many path breaking experiments were conducted during the twentieth century which established the nature of the gene and the chromosome, most important being identification of DNA as the genetic material by the classical Alfred Hershey and Martha Chase experiment in 1952. James Watson and Francis Crick proposed the double helical structure of DNA in 1953. Many experiments followed that determined how the information in the gene is used, such as the manipulation of enzymes involved in DNA replication, and DNA repair.

Modern biotechnology is based on the rDNA technology that has revolutionised biotechnology by allowing scientists to cut and join different pieces of DNA, and place the new recombinant (Chimeric/ hybrid) DNA into a new host (Fig. 1.1). It allows the transfer of gene(s) from one organism to another conferring a novel property. This has revolutionised the age old process of biotechnology with regards to its precision and efficiency with limitless possibilities. Since the advent of rDNA technology, biotechnology has become more advanced and led to advancements in medicine, agriculture, animal science and environmental science. The multi-disciplinary nature of modern biotechnology and the areas of its application is given in Fig. 1.2 and Table 1.1.

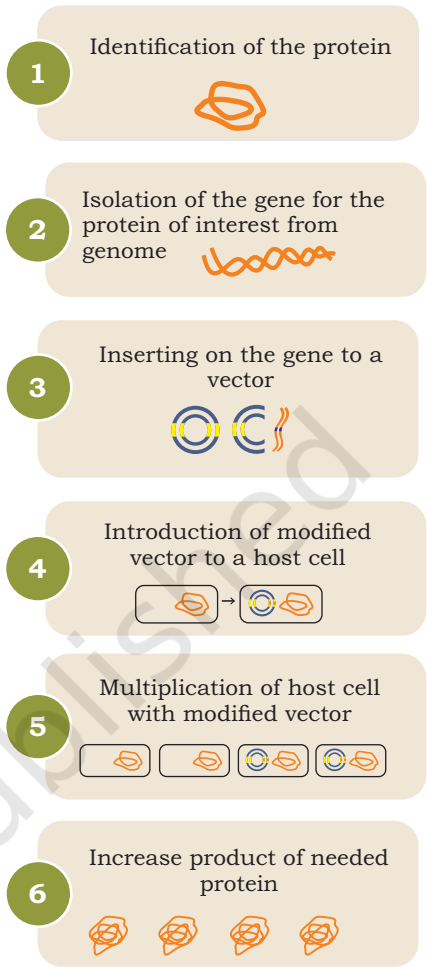


Fig. 1.1: Overview of modern biotechnology

Table 1.1: Some common names of areas covered under biotechnology

Blue Biotechnology	Application of biotechnology for marine and freshwater organisms, which are used for increasing seafood supply, regulation of the reproduction of dangerous water-borne organisms, and developing new drugs.
Green Biotechnology	Application of biotechnology for environment-friendly solutions such as in plants to improve the nutritional quality, quantity and production of eco-friendly products. The transgenic plants with improved traits are the examples of green biotechnology.

© NCERT not to be republished

Red Biotechnology	Medical biotechnology which is applied to manufacture pharmaceutical products such as insulin, enzymes, antibiotics and vaccines.
White Biotechnology	Biotechnology applied to improve industrial processes and other production processes. The use of enzymes as industrial catalysts in eco-friendly manner for the production of valuable chemicals.

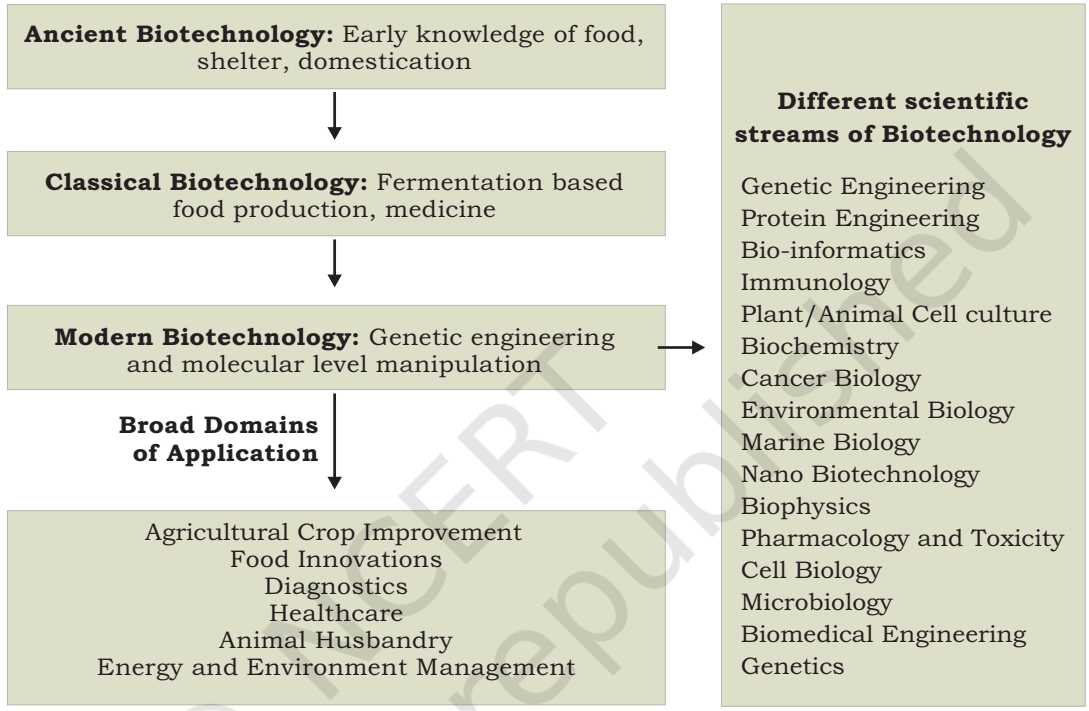


Fig. 1.2: Multi disciplinary nature of modern biotechnology and the areas of its application

1.2 APPLICATIONS OF MODERN BIOTECHNOLOGY

Modern biotechnology, which is based on rDNA technology, exhibits a wide range of applications. The broad application areas of biotechnology include pharmaceutical and therapeutic research, disease diagnostics, crop improvement, vegetable oil, biofuels, and development of enviromental friendly products (for example biodegradable plastics). Some classic examples of successful application of biotechnology is provided in Fig. 1.3. Thus, the applications of modern biotechnology mainly focus on the following major areas:

1. Medicine and health care
2. Crop production and agriculture
3. Food processing
4. Environmental protection

© NCERT not to be republished

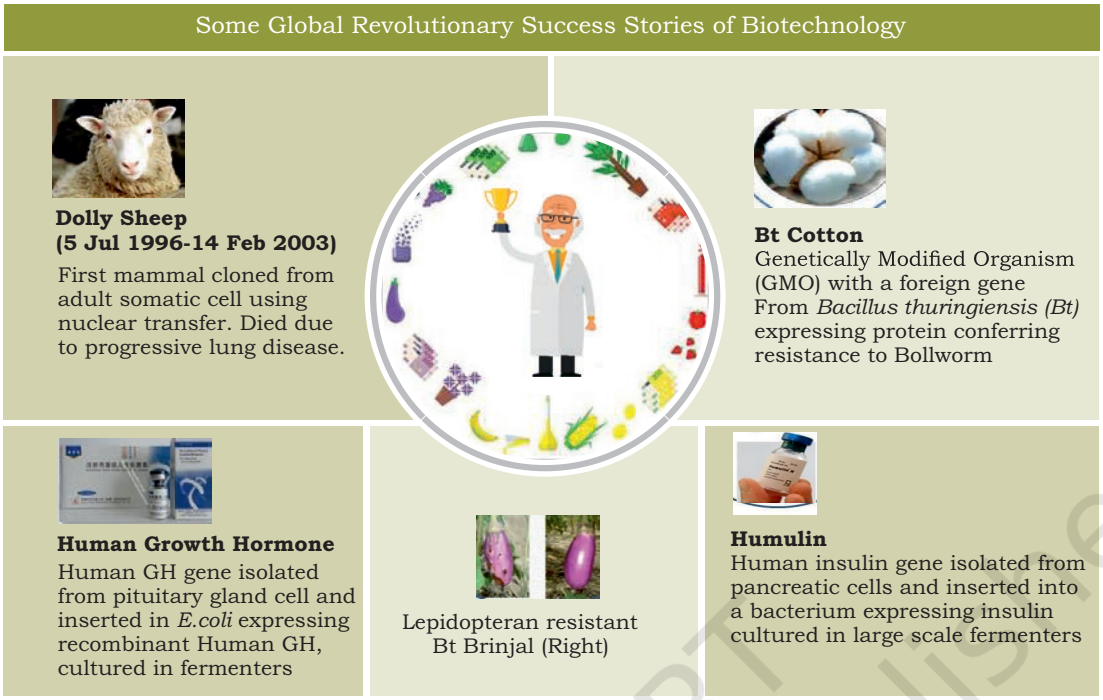


Fig. 1.3: Some classic examples of successful applications of biotechnology

1.2.1 Medicine and health care

Biotechnology techniques are used in the field of medicine for diagnosis via development of diagnostic tools and kits, which have proved helpful in detecting certain molecules and cellular components which are expressed in diseased conditions. Using rDNA technology, tools of bioinformatics, modern instrumentation and bioprocess technologies, synthetic drug analogs can be predicted and possibly synthesised, which may show improved disease treatment. Production of vaccines and gene therapy are also important applications of biotechnology in the field of medicine.
Some of the major applications of modern biotechnology in the field of medicine are listed below:

- **Production of important therapeutic molecules:** rDNA technology has been successfully applied for the development of biopharmaceuticals with therapeutic value. Different protein molecules which may act as drug molecules, are being expressed in heterologous systems such as microorganisms, plants (transgenic plants explained in the following section) etc.

  A number of therapeutic products including antibiotics and hormones have been produced using

© NCERT not to be republished

rDNA technology which are available in the market. A common example of therapeutic protein produced using rDNA technology is human insulin, used for the treatment of diabetes, a disease in which blood sugar levels are elevated. This presents a classic example of a human protein being expressed in a heterologous system such as *Escherichia coli.* At present, insulin is being produced predominantly in *E. coli* and *Saccharomyces cerevisiae*. The human growth hormone is another example of successful production of desired proteins in different microbial host systems via rDNA technology. Many human proteins have also been expressed in milk of transgenic sheep and goat. For example, Food and Drug Administration, USA (FDA) has approved the production of blood anti-coagulant in milk of transgenic goats for human use.

Currently, scientists are trying to develop such drugs against diseases like hepatitis, cancer and heart diseases, which are the leading causes of human mortality.

- **Gene therapy:** This technology is most helpful in the treatment of diseases caused by gene defects such as cystic fibrosis, thalassemia, Parkinson's disease, etc. Conceptualised in 1972, gene therapy involves delivery of required gene into a patient's cell as a drug to treat disease, so that it replaces the function of the defective gene. The first attempt, although unsuccessful, was performed by Martin Cline in 1980 for treating β-thalassemia. The first successful report of gene therapy was achieved in 1990 when, Ashanthi De Silva was treated for Adenosine Deaminase deficiency [also called Adenosine Deaminase Severe Combined Immunodeficiency (ADA-SCID)] which is an autosomal recessive metabolic disorder that causes immunodeficiency. Russia approved Neovasculgen in 2011, as a first-in-class-gene therapy for peripheral artery disease.

The first commercial gene therapy product approved for cancer treatment by China in 2003, was **Gendicine**.

- **Genetic testing:** It is a type of medical test that helps in identifying the defects in an individual's genetic composition such as chromosomal defects in gene and protein expression anomalies. It helps in determining a person's chance of developing or passing on a specific disorder. Hundreds of genetic tests are currently in use and many are being developed. For example, genetic

© NCERT not to be republished

tests for phenylketonuria (patients lack enzyme needed to metabolise the amino acid phenylalanine, which is responsible for normal growth) and congenital hypothyroidism (thyroid gland disorder) have been developed.

1.2.2 Crop Production and Agriculture

Biotechnology has played a major role in revolutionising agriculture by facilitating genetic manipulations of important crop plants, to develop biotic and abiotic stress resistance plants, and better quality products of food in terms of nutrition and longer shelf lives. The five major traits used for crop improvement are insect resistance, herbicide resistance, virus resistance, delayed fruit ripening and nutritional enhancement. Thus, transgenic plants (Genetically Modified Organisms; GMOs) harbouring these improved traits are good examples of the application of biotechnology in agriculture.

Some examples describing the success stories of biotechnology are given below:

Biotechnology for crop improvement

- Biotechnology, based on rDNA technology, has immense applications in crop improvement. Although conventional plant breeding techniques have made considerable progress in the development of improved varieties, they have not been able to keep pace with the increasing demand for food, vegetables and fruits. Use of rDNA technology has successfully led to the development of a number of transgenic plants exhibiting resistance to pathogens, salt, cold, herbicide, etc. In these transgenic plants, useful genes have been stably incorporated into the plant genome which resulted in the stable expression of targeted gene product.
- Among biotic stress resistant category, for example, virus-resistant plants have a viral coat protein gene which is overproduced that prevents the virus from reproducing in the host cell. Coat protein genes are involved in resistance against many viruses such as Papaya Ring Spot Virus, Cucumber Mosaic Virus, Tobacco Rattle Virus, and Potato Virus in these plants.
- Crop losses from insect pests also result in devastating financial loss of farmers and may lead to starvation in developing countries. Spraying chemical pesticides

© NCERT not to be republished

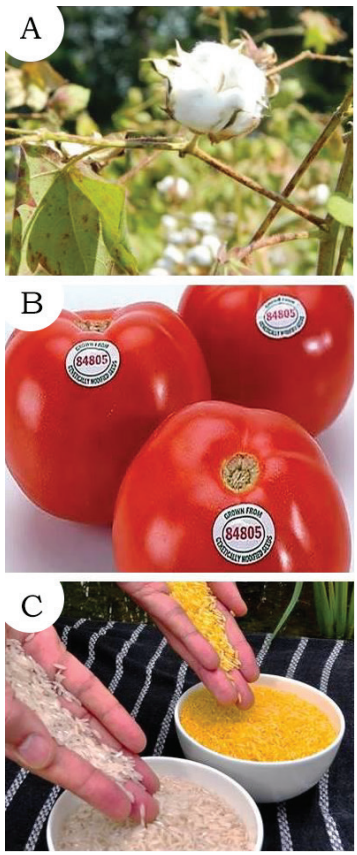


Fig . 1.4: Examples of some GM plants: (A) Bt Cotton (B) Flavr Savr Tomato and (C) Golden rice

is costly and it leads to potential health hazards and may also pollute the environment. Genetically modified plants offering resistance to insect pathogen have been developed through rDNA technology. These plants help in bringing down or eliminating the application of chemical pesticides. One common example is Bt cotton. Bt is a toxic protein called Cry 1A(b), obtained from a soil bacterium called *Bacillus thuringiensis*, demonstrates insecticidal activity against larvae of moths and butterflies, beetles, cotton bollworms and caterpillars but are harmless to us. Thus, the gene coding for Bt toxin has been transferred and expressed in cotton. These transgenic cotton plants express Bt toxin which acts as insecticide [Fig. 1.4(A)]. Similar to the Bt cotton, other plants including brinjal, corn (maize), potato, soybean, tomato, tobacco have also been developed expressing the Bt toxin.

- Among abiotic stress resistant plants, resistance against chilling has been introduced into tobacco plants by introducing gene for glycerol-1-phosphate acyl-transferase enzyme from *Arabidopsis.* Similarly, Roundup-ready soybeans (Transgenic/GM soybeans) have been developed which are unaffected by the herbicide glyphosate, and therefore can be applied in selective killing of competing weeds.
- Among the quality improvement category, a classic example is the development of Flavr Savr tomato. These tomatoes have extended shelf life due to delayed ripening [Fig. 1.4(B)].
- Biotechnological tools have also been extensively used to improve the nutritional quality of different food crops. A classic example is the Golden Rice, which has high beta-carotene content (the precursor for vitamin A production in the human body) [Fig. 1.4(C)]. The name comes from the colour of the transgenic grain due to over expression of beta-carotene, responsible for golden coloration.
- The technique of plant tissue culture, i.e., culturing plant cells or tissues in artificial medium supplemented with required nutrients, has many applications in efficient clonal propagation (true to the type or similar) which may be difficult via conventional breeding methods. Many of the dry land legume species have been successfully regenerated from culture of

© NCERT not to be republished

cotyledons, hypocotyls, leaf, ovary, protoplast, petiole root, anthers, etc. Haploid generation through anther/ pollen culture is recognised as another important area in crop improvement. Storage of horticultural crops with recalcitrant seeds or perennial crops may be maintained via plant cell culture, which is of great practical importance. These techniques have successfully been demonstrated in a number of horticultural crops and now there are various germplasm collection centers globally.

Transgenic plants as systems for expression of therapeutics

Plants can also be used as heterologous systems for expression of a therapeutic molecule via expressing the required gene(s) into the plant using rDNA technology. An example is the production of antibiotics particularly for animal use in stock feed plants. Stock feed plants are plant species that may be given to cattle and livestock as a food source. Examples of such stock feed are bamboo, citronella, andropogon, foxtail millet, wheat grass, rice straw, etc. Stockfeeds capable of stably expressing the desired antibiotic may be fed directly to animals. This technique is less expensive than traditional antibiotic production and administration. However, this practice raises many bioethical issues, especially in the arena of human use, because of possible development of drug resistant bacterial strains due to antibiotic overuse.

Similarly, transgenic plants have been developed for production of **edible vaccines** by expressing antigenic proteins from pathogens into the edible parts of the plant, in a form that will retain its immunogenicity. Individuals are expected to be immunised by simply consuming such transgenic plants. Potato based vaccines against measles, cholera, Norfolk virus, etc., are under rigorous clinical trials.

Biofuels

These production can also be improved using biotechnology. These are produced through biological processes rather than a fuel produced by geological processes such as coal and petroleum. Biofuels can be derived directly from plants, or indirectly from agricultural, commercial and industrial wastes. Basically, it involves generation

© NCERT not to be republished

of biomass that can be converted to convenient energy-containing substances via different ways such as thermal conversion, chemical conversion, and biochemical conversion. This biomass conversion can result in fuels which are in solid, liquid, or gas form. Major types are bioethanol or biologically generated alcohols produced via fermentation of sugar and starches by micro-organisms. **Bio-butanol**, a biofuel, is often a direct replacement for gasoline. Biodiesel is the most common biofuel in Europe, produced from oils or fats using trans-esterification. Feed stocks for biodiesel include animal fats, vegetable oils, soy, rapeseed, *Jatropha*, hemp, etc. Other examples are **bio-ethers** and **biogas**.

The first commercial-scale plants to produce biofuels from cellulose containing organic matter, have begun operating in the United States. In parts of Asia and Africa where drylands prevail, sweet sorghum is being investigated as a potential source of food, feed and fuel. Since the crop uses very little water, it is particularly suitable for growing in arid conditions. In India, and other places, sweet sorghum stalks are used to produce biofuel by squeezing the juice and then fermenting into ethanol. Several groups in various sectors are conducting research on ***Jatropha curcas***, which produces seeds considered to be a viable source of biofuels feedstock oil. Current research focuses on improving the overall oil yield of *Jatropha* through biotechnological techniques.

1.2.3 Food processing

The role of biotechnology in food processing is immense as discussed below:

- Biotechnological tools can help in improving the edibility, texture, and storage of the food; prevention of mycotoxin production, extending shelf life and also to delay time dependent degradation of nutritional components of foodstuffs.
- Almost one-third of the world's diet consists of fermented food.
- Protein engineering of microbial enzymes capable of improved fermentation are produced commercially at a large scale by culturing the microorganisms in tanks and industrial scale fermenters.

© NCERT not to be republished

- Industrial scale production of fermented foods with added taste, nutrition and shelf life such as cheese, yoghurt, certain probiotics, buttermilk and other popular fermented products has also been made possible.

1.2.4 Environmental protection

Biotechnological tools and techniques are also very helpful in tackling issues related to environment and ecology. A special branch of science which applies biotechnology to study the natural environment, identifying optimum, but sustainable uses of plants, animals and micro-organisms to develop green technology, and remediation of contaminated environments is known as Environmental Biotechnology. Some of the remarkable achievements obtained in environmental biotechnology are as follows:

- Many eco-toxicological biomarkers are being used to indicate the effect of xenobiotics which are present in the environment as well as within an organism. Bio-markers are defined as any naturally occurring molecule which may indicate specific biological processes in response to any environmental or chemical stimuli. Many eco-toxicological biomarkers are developed which may prove helpful in indicating subtle changes in the immediate environment, which may otherwise be difficult to detect. For example, a reported gene, *lux* (which is responsible for emission of light), expressed in *E. coli*, acts as a bio sensor for detecting the mercury contamination.
- Biotechnological applications may also be helpful in the process of cleaning up the hazardous substances in the environment by converting them into non-toxic or less toxic compounds. This is known as **bioremediation**. This process of clean-up exploits the potential of natural sources for bioremediation. Genetic engineering has been exploited to generate organisms specifically designed for bioremediation. Genes, which code for enzymes for degradation of pollutants or monitor their levels may be inserted into the organisms. An example of a degradation gene is biphenyl dioxygenase, which has been inserted in *E.coli* to degrade PCB (polychlorinated biphenyl).
- Cultivable land area is often contaminated with heavy metals such as Cadmium, Mercury and Lead, which

© NCERT not to be republished

may prove detrimental for growth of crop plants and may even prove as a health hazard upon consumption. Many hyper-accumulator plants when grown on these contaminated soils, have the potential to soak up the heavy metals from the soil and sequester it in their cellular compartments, thereby phyto-remediating the soil. Examples of some hyper-accumulators are ***Brassica napus and Helianthus annus,*** etc., for mercury and lead removal from contaminated soils. Extensive research is being carried out in identifying the genes responsible for tolerance of these plants to such hazardous heavy metals.

- The application of environmental biotechnology will help to keep our environment safe and clean for future generations. It can provide alternative ways of adaptation to the changes in the environment. Multi-disciplinary association between branches of science such as genomics, proteomics, bioinformatics, sequencing and imaging processes provide large amounts of information and novel ways to protect the environment.

1.3 BIOTECHNOLOGY IN INDIA: ACADEMIC PROSPECTS AND INDUSTRIAL SCENARIO

Few of the first biotechnology firms to be set up in India were the Serum Institute of India (late 1960s) and Biocon (1978) recognising the significance of biotechnology. The National Biotechnology Board (NBTB) was constituted by Government of India in 1982, which was subsequently upgraded to Department of Biotechnology (DBT) in 1986. DBT has established some major research institutions all over India (Table 1.2).

DBT also supports post graduate academic programmes of biotechnology in the subject areas of Agricultural Biotechnology, Marine Biotechnology, Neuroscience, Industrial Biotechnology, Environmental Biotechnology, Bioresources, etc. These courses are being conducted at different centres located at various State and Central Universities of India. Certificate and Diploma courses in biotechnology are also offered by some premier Indian institutes.

Emphasis for development of human resource was one of the mandates of DBT. DBT offers many research fellowships for promising students interested in the area of biotechnological research. Among these are DBT scholarships after schooling, DBT-JRF to support doctoral research in biotechnology, DBT research associateship (DBT-RA) to support life sciences post-doctoral study in the premier institutes of India.

© NCERT not to be republished

Table 1.2: List of institutes established under DBT engaged in active research

S.No.	Name of the Institutes
1.	Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, Telangana
2.	Institute of Bioresources and Sustainable Development (IBSD), Imphal, Manipur
3.	Institute of Life Sciences, Bhubaneswar, Odisha
4.	National Agri-food Biotechnology Institution (NABI), Mohali, Punjab
5.	National Brain Research Centre (NBRC), Gurugram, Haryana
6.	National Center for Cell Science, Pune, Maharashtra
7.	National Institute for Plant Genome Research (NIPGR), New Delhi
8.	National Institute of Animal Biotechnology (NIAB), Hyderabad, Telangana
9.	National Institute of Biomedical Genomics (NIBMG), Kalyani, West Bengal
10.	National Institute of Immunology (NII), New Delhi
11.	Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram, Kerala
12.	Regional Centre for Biotechnology (RCB), Faridabad, Haryana
13.	Institute of Stem Cell Science and Regenerative Medicines, Bengaluru, Karnataka
14.	Translational Health Science and Technology Institute, Faridabad, Haryana

© NCERT not to be republished

1.3.1 Indian biotechnology industry

The Indian biotechnology industry is one of the fastest growing industries of the country. Presently, India is among top 12 biotechnology powers in the world and third biggest industry in Asia Pacific in terms of industrial biotechnology infrastructure.

The Indian biotechnology industry is divided into five segments namely, Bio-pharma, Bio-services, Bio-agri, Bio-industrial, and Bio-informatics. India has also obtained significant achievement in Bio-agri segment. India has fourth largest area covered by genetically modified crops.

The Bio-pharma industry comprises mainly of vaccines manufacturing and its export in large quantities internationally. The other products that are being

produced at large scale are diagnostics and therapeutics. There are many Biopharma company in the country which contributes to a large number of products related to pharmaceuticals and other medicine produts. Indian companies hold expertise to indigenously develop and manufacture many recombinant biotech products such as recombinant Hepatitis B Vaccine, Human Insulin, G-CSF, Erythropoietin, Human Growth Hormone and Interferon alpha 2b. India is world's largest producer of recombinant Hepatitis B vaccine. Bharat Biotech commercially launched the first ROTAVAC® vaccine to eradicate rotavirus diarrhoea in India. The Government of India also initiated collaborations with private pharmaceutical companies for better Research and Development (R and D). BIRAC (DBT) and Department of Electronics and Information Technology (DeitY) collaborated with each other in 2016 to promote innovative technologies in medical electronics sector.

Biotechnology is an industrial hub and is refining the economy of the nation. The ten most recognised Biotech companies are listed in Table 1.3. Various biotechnology parks (Table 1.4) have also been established in collaboration with these firms and the Government of India to provide high end research infrastructures.

1.3.2 Achievements and innovations

Vaccines

Major products of bio-pharma industry are vaccines and therapeutics. India is the world's largest producer of recombinant Hepatitis B vaccine, measles vaccine and DTP vaccine. World's only adsorbed liquid HDC rabies vaccine and India's first MMR vaccine 'Tresivac' has been launched by the Serum Institute. One out of every two children in the world has been vaccinated by vaccines manufactured by an Indian company. The country exports its vaccines to more than 140 countries as well as to UNICEF and the Pan American Health Organization (PAHO). Major vaccine producing firms of the country are Serum Institute of India Ltd., Bharat Serum and Vaccines Limited, Panacea Biotec, Glaxo Smith Kline Pharmaceuticals Limited, Wockhardt Biotech, etc. with international repute.

© NCERT not to be republished

Table 1.3: List of a few biotech companies functional in India (in alphabetical order)

Bharat Biotech International Limited Company
Bharat Serum and Vaccines Limited
Biocon
Dr. Reddy's Laboratories
GlaxoSmithKline Pharmaceuticals Limited
Indian Immunologicals
Novozymes
Panacea Biotec
Serum Institute of India Limited
Shantha Biotechnics Limited (Sanofi)
Wockhardt Biotech Park

Table 1.4: Few biotechnology parks in India (in alphabetical order)

Parks	City
Bangalore Biotech Park	Bengaluru, Karnataka
Bio-Pharma-IT Park	Bhubaneswar, Odisha
Golden Jubilee Biotech Park	Chennai, Tamil Nadu
Guwahati Biotech Park	Guwahati, Assam
ICICI Knowledge Park	Hyderabad, Telangana
International Biotech Park	Pune, Maharashtra
KINFERA Biotech Park	Kochi, Kerala
Lucknow Biotech Park	Lucknow, Uttar Pradesh
Shapoorji Pallonji Biotech Park	Hyderabad, Telangana
Ticel Bio Park	Chennai, Tamil Nadu

In 2016, Memorandum of Understanding (MOU) was signed by Sun Pharmaceutical Industries Ltd. and International Center for Genetic Engineering and Biotechnology (ICGEB), India, for the development of vaccine for all forms of serotypes of the dengue virus. India is also amongst the first few countries that developed vaccines against Covid-19. Covaxin, India's first indigenous Covid-19 Vaccine, was developed by a Biotechnology Company, Bharat Biotech International Ltd. Hyderabad (Telangana).

© NCERT not to be republished

Therapeutics

Stem cell research, monoclonal antibody products, growth factors, cell engineering and cell based therapeutics are other areas of the bio-pharmaceutical industry. India is the world's largest producer of Statin and Immunosuppressants. In 2015, launched INSUPEN, a convenient and affordable reusable insulin delivery device. In 2016, the first genuine formulation Rosuvastatin tablet was granted European approval and was launched in 2017. Panacea Biotec has been selected by WHO for developing the sabin based injectable polio vaccine. 80 per cent of the antiretroviral drugs for AIDS are provided by Indian pharmaceutical firms globally.

Agriculture sector

Agriculture sector has immense opportunities in India. Indian hybrid seed industry is growing at the rate of 10–17 per cent annually with Bt cotton leading the market. Monsanto Research Centre for plant genomics was established in 1998 in India. In the same year DBT approved Monsanto to grow Bt cotton in India. Since then Bt cotton is leading the market, accounting for 45% share in Indian hybrid seed industry. India has the potential to become a major producer of transgenic rice and several GM (genetically modified) crops and vegetable hybrid seeds, including GM seeds. Among these GM rice is the Blight resistant rice, *Samba Mahsuri* (developed through marker assisted backcrossing), flood, drought and salt tolerant rice, etc. GM maize hybrid with high quantity of protein and increased provitamin A has also been developed. Bread wheat and durum wheat genotypes with high yield and high micronutrient concentration are also being developed.

Bio-services

Bio-services sector represent an area of significant promise for India because of a huge skilled labour force, attractive cost and access to major markets in Asia. It includes global contract research organisations, such as Quintiles, as well as Indian companies including GVK Bio, Jubilant Biosys and Advinus.

© NCERT not to be republished

SUMMARY

- Biotechnology refers to the different technologies that make use of living cells and/or biological molecules to generate useful products for the benefit of mankind.
- The beginning of cultivation of crops as early as in the Paleolithic Era, around 10,000 years ago, marks the introduction of biotechnology and is commonly referred to as ancient biotechnology.
- Production of bread, cheese, wine, vinegar, beer, etc., is the contribution of ancient biotechnology.
- Modern biotechnology is based on recombinant DNA technology which allows scientists to cut and join different pieces of DNA and place the new recombinant DNA into a new host allowing the transfer of gene(s) from one organism to another conferring a novel property.
- Modern biotechnology has its applications in numerous fields such as Agricultural Crop Improvement, Food Innovations, Diagnostics Healthcare, Animal Husbandry, and Energy and Environment Management.
- The different streams of modern biotechnology are Genetic Engineering, Protein Engineering, Bio-informatics, Immunology, Plant/Animal cell culture, Biochemistry, Cancer Biology, Environmental Biology, Marine Biology, Nano Biotechnology, Bio-physics, Pharmacology and Toxicity, Cell Biology, Microbiology, Biomedical Engineering, and Genetics.
- The Government of India constituted the National Biotechnology Board in 1982 which was subsequently upgraded to Department of Biotechnology (DBT) in 1986. DBT has established several major research institutions all over India.
- Indian biotechnology industry is one of the fastest growing industries of the country, presently ranked among the top 12 biotechnology powers in the world and third biggest industry in Asia Pacific in terms of industrial biotechnology infrastructure.
- Indian biotechnology industry has made significant contributions in the field of Vaccine production, Therapeutics, Agriculture sector, and Bio-services.

© NCERT not to be republished

EXERCISES

1. What do you understand by the term 'Biotechnology'? Explain giving suitable examples.
2. Give a comparative account of the ancient and modern concept of biotechnology.
3. Elaborate on the role of biotechnology with respect to the following:
   (a) Biopharmaceutical production
   (b) Gene therapy and applications
   (c) Abiotic stress resistance in crops
   (d) Crops with insect resistance
   (e) Environmental protection and conservation
4. Explain the contribution of ancient biotechnology in human welfare.
5. Modern biotechnology is based on recombinant DNA technology. Justify the statement.

© NCERT not to be republished

Chapter 2
Cellular Organelles

Chapter 3
Biomolecules

Chapter 4
Enzymes and Bioenergetics

Chapter 5
Cellular Processes

Unit II
Cell Organelles and Biomolecules

Being the structural and functional unit of a living organism, cell has got a very important place in understanding the entire functioning of a living system. Therefore, it is required to have a thorough understanding of the structure and functions of a cell. This unit gives a detailed description of the general characteristics of cell, its structure and growth. Cell theory, which will be explained in Chapter 2 of this unit, offered an intriguing explanation of the living phenomena. It filled the researchers with wonder and compelled them to investigate the molecular structure and physiochemical characteristics of a cell. Chapter 3 deals with details of the molecular compounds responsible for carrying out various cellular processes and sustaining living systems. Central to the various molecular processes are the biocatalysts, which help in carrying out various biochemical reactions in a living system effectively. The importance of biocatalysts and their mechanism of action will be discussed in Chapter 4. The concepts and mechanisms of some of the important cellular processes have been dealt with in Chapter 5.

© NCERT not to be republished

G.N. Ramachandran (1922–2001)

He was born in a small town near Cochin on the Southwestern coast of India. His father taught mathematics at a local college and profoundly influenced Ramachandran's interest in mathematics. Ramachandran completed his graduation in physics in the year 1942 and was the top-ranking student at his college. He received a doctoral degree in 1949 from Cambridge University. His meeting with Linus Pauling considerably directed his attention towards solving the intricate structure of collagen. In 1954, his study on the Triple Helical Structure of Collagen was published in the *Nature*. Ramachandran was the founder of 'Madras School' of conformational analysis of biopolymers. His work on the analysis of the allowed conformations of proteins through the use of 'Ramachandran plot' is considered to be one of the most outstanding contributions in structural biology.

© NCERT not to be republished

11150CH02

Chapter 2

Cellular Organelles

2.1 *Plasma Membrane*
2.2 *Cell Wall*
2.3 *Endomembrane System*
2.4 *Mitochondria*
2.5 *Plastids*
2.6 *Ribosomes*
2.7 *Microbodies*
2.8 *Cytoskeleton*
2.9 *Cilia and Flagella*
2.10 *Centrosome and Centrioles*
2.11 *Nucleus*
2.12 *Chromosome*

Overview

Our body does large number of tasks at a given point of time for example, food digestion, sending electrical messages through nerves, pumping blood from the heart, circulating nutrients, synthesising proteins, filtering urine and many more. All this is possible because of cells which are considered to be the basic unit of life. Each cell is equipped with different machineries known as **organelles** responsible for different functions. You also know that cells present in organisms (unicellular or multicellular) have been broadly characterised into two main categories, based on the nuclear organisation and membrane-bound cell organelles, i.e., prokaryote and eukaryote. Some of the components are common to both prokaryotic and eukaryotic cell. These are plasma membrane, cytoplasm, ribosomes, DNA, etc. Prokaryotic cells are without an organised nucleus and contain numerous ribosomes, mesosomes (folds in plasma membrane) besides having locomotory structures such as flagella in some of them. While a eukaryotic cell has a well-organised nucleus, cell

© NCERT not to be republished

membrane and membrane-bound cell organelles such as endoplasmic reticulum, Golgi apparatus, mitochondria, plastids, vacuole, lysosomes, peroxisomes, and many more. Advancement in microscopic techniques played a very crucial role in exploring the detailed structure of cell.

Let us now look at an individual cell to understand the structure and functioning, along with the role in establishing cell functioning and life.

2.1 PLASMA MEMBRANE

Plasma membrane forms the boundary of the cytoplasm being guarded from outside by extracellular matrix. The membrane is responsible for the relationship of a cell with its surrounding. It is semipermeable in nature. Major breakthrough in understanding the detailed structure of cell membrane was realised only after understanding the chemical composition (lipid and protein mainly) and the invention of electron microscope. A widely accepted model for the organisation of plasma membrane was proposed by Seymour Jonathan Singer and Garth L. Nicolson (1972) as '**The Fluid Mosaic Model**' (Fig. 2.1). The model suggests plasma membrane to be lipid bilayer surrounding the cell with mosaic of globular proteins. Composition of lipid and protein varies in different cells, for example, human erythrocyte membrane contains approximately 52 per cent protein and 40 per cent lipids. Lipid bilayer makes the cell boundary in a quasifluid state and it is dynamic in nature. Due to the fluid nature, lipids and proteins can freely diffuse laterally across the membrane. Phospholipids (the major membrane lipid) is composed of hydrophilic head facing the exterior and long hydrophobic tail of hydrocarbon chains occupying the interior of a lipid bilayer. Two different types of proteins have been identified in the plasma membrane based on their location and association i.e., **peripheral** and **integral membrane proteins**. Peripheral membrane proteins are mainly involved in cell signalling and these are superficially attached to lipid bilayer. Integral membrane proteins are partially or fully buried in the plasma membrane. Transmembrane proteins are the most abundant type of integral membrane protein. Structurally, prokaryotic cell membrane is similar to that of eukaryotes.

© NCERT not to be republished

Box 1

Edwin Gorter and F. Grendel in the year 1925 collected blood cells (chromocytes) from the artery or vein of mammals. The chromocytes were separated from plasma by several washes with saline solution and extracted using acetone. They obtained lipids that exactly covered the entire surface area of chromocytes like a two-molecular thick layer. They observed all the cells, either prokaryotic or eukaryotic, to be enclosed with well-defined plasma membrane, which maintains cell identity by preserving its internal constituents from the environment. This evidence was further supported by high magnification electron micrograph referring plasma membrane as a 'railroad track', with two densely stained lines of polar heads groups of phospholipids and a lightly stained portion representing hydrophobic fatty acid chain. Its molecular organisation was still rudimentary. On the basis of this, they proposed the bilayer structure of plasma membrane rather than a monolayer, using mammalian RBCs as a model.

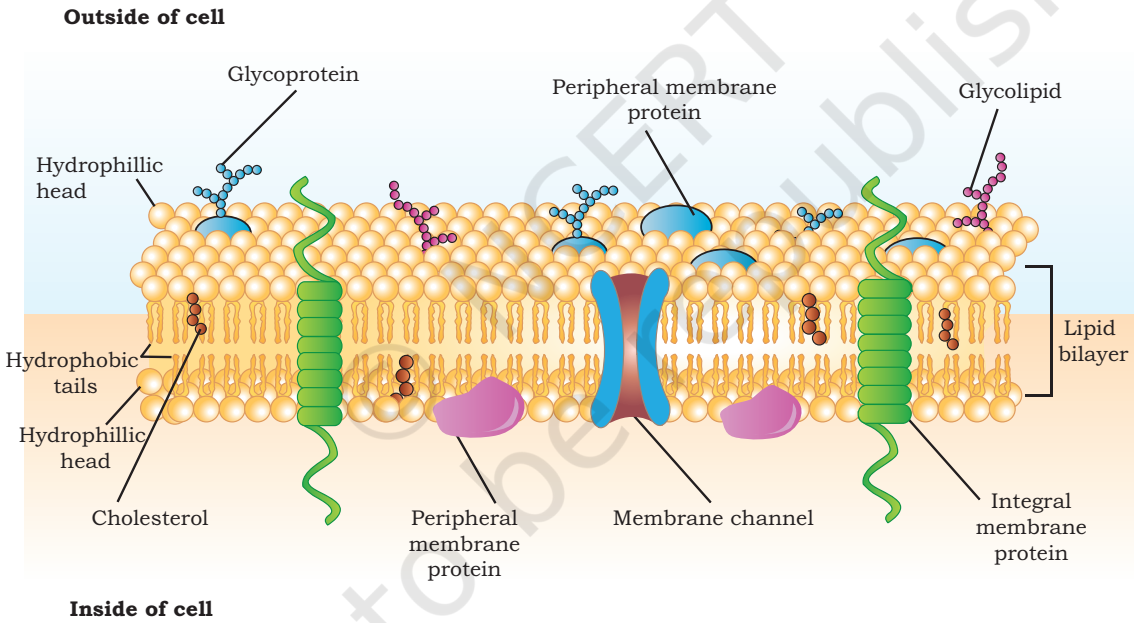


Fig. 2.1: Schematic diagram showing fluid mosaic model of plasma membrane

A special membranous structure is formed by extension of plasma membrane in the cell, this structure called **mesosome**, which is in the form of vesicle, tubules and lamellae. Mesosomes increase the surface of plasma membrane.

The quasifluid nature of membrane is useful for different cellular functions such as cell division, cell growth,

© NCERT not to be republished

communication at intercellular junctions, cell secretion, endocytosis, etc. Plasma membrane being selectively permeable restricts molecular movement and maintains cell composition. Some of the molecules move passively without any expenditure of energy across the membrane along the concentration gradient called **passive transport**. Passive movement of molecules occurs by the process of diffusion and osmosis. However, a few molecules either charged (for example, ions and amino acids) or uncharged (for example, glucose) cannot cross plasma membrane by simple diffusion. Movement of such molecules is facilitated by **carrier proteins** for example, glucose transporter and **channel proteins** [Fig. 2.2 (a,b)]. Such molecular movement is known as the **facilitated movement**. **Aquaporins** are one of the critical channel proteins for transport of water in plant and animal cell across the plasma membrane. Some of the well-studied channel proteins in the membrane of muscle and nerve cell are **ion channels** [Fig. 2.2(b)].

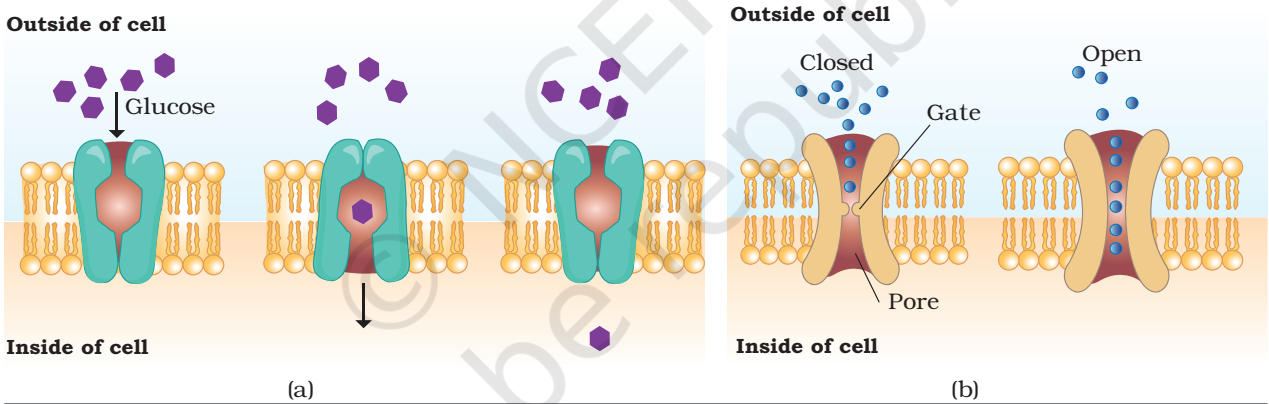


Fig. 2.2: Membrane transport (a) Facilitated transport of glucose and (b) Transport through ion-gated channel

Molecules which are transported against the concentration gradient (i.e., from lower concentration to higher concentration) require utilisation of energy from ATP molecules, e.g. **Na^+-K^+ pump** (Fig 2.3). This is referred to as **active transport**. However, some active transports are ATP-independent; molecules are transported against the concentration gradients with no energy utilisation from ATP hydrolysis. It couples transport of such molecule with a second molecule transported along the concentration gradient for example, active transport of ions, sugars and

© NCERT not to be republished

amino acids using energy derived from the Na^+ gradient.

In coupled transport, if two molecules are transported in the same direction (uptake of glucose and Na^+), it is called **symport**. If active transport involves transport of two molecules in the opposite direction (transport of Na^+ and Ca^{2+} by Na^+-Ca^{2+} antiporter), it is called **antiport**. While facilitated diffusion transports only single molecule, for example, glucose, it is known as **uniport**.

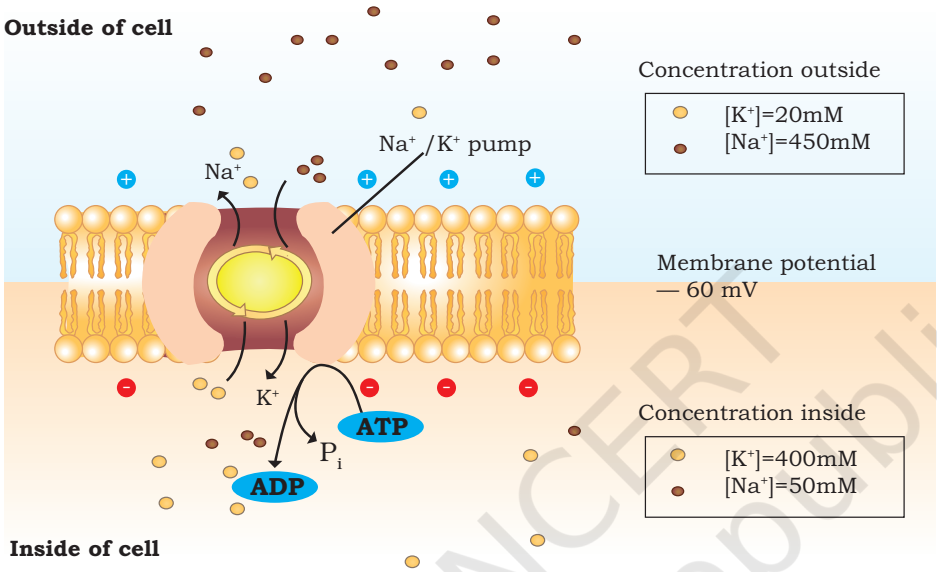


Fig. 2.3: Active transport through Na^+ - K^+ pump

2.2 CELL WALL

Cells of bacteria, algae, fungi, and higher plants are also surrounded by a rigid cell wall besides the plasma membrane. It is not found in animal cells. It differs structurally in bacteria and eukaryotes. In bacteria, it is composed of polysaccharide cross-linked by small peptides, which provides rigidity, shape and protection from osmotic pressure; and in eukaryotes (plants and fungi), it is primarily made up of polysaccharides. Cell wall not only determines the cell shape, but also prevents cell bursting caused due to osmotic pressure. It also helps in cell–cell interaction and provides mechanical strength and protection from infection. Gram-positive bacteria have a thick cell wall with the single plasma membrane [Fig. 2.4 (a)]. In contrast, Gram negative bacteria have a thin cell

© NCERT not to be republished

wall surrounded by a dual plasma membrane [Fig. 2.4 (b)]. Cell wall continuously grows and changes its shape as bacteria grows and divides. Structurally, the bacterial cell wall is a sturdy covalent shell of linear peptidoglycan chain cross-linked by tetrapeptides. Commonly used antibiotics are known to inhibit this cross-linking of peptidoglycan strands and interfering bacterial growth.

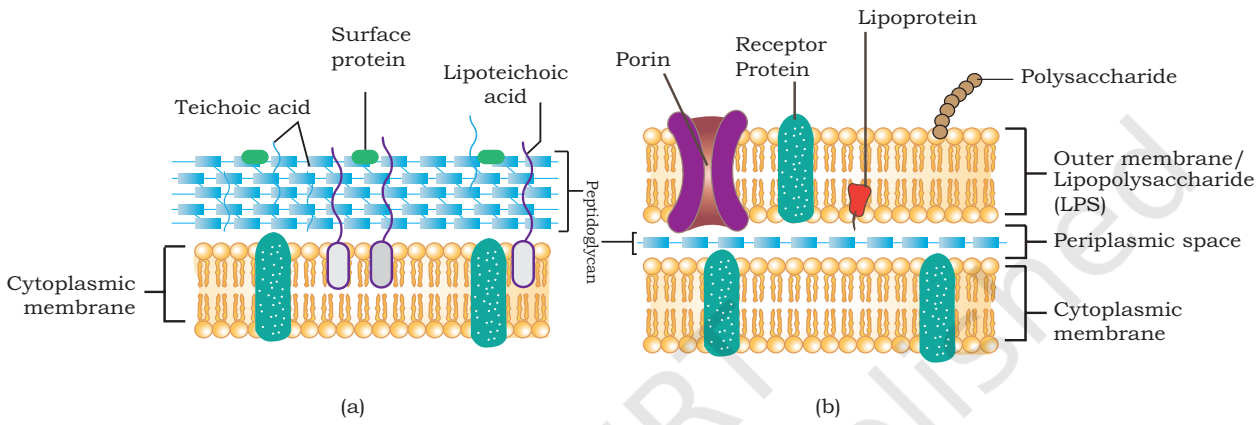


Fig. 2.4: Prokaryotic cell wall; (a) Gram-positive and (b) Gram-negative bacteria

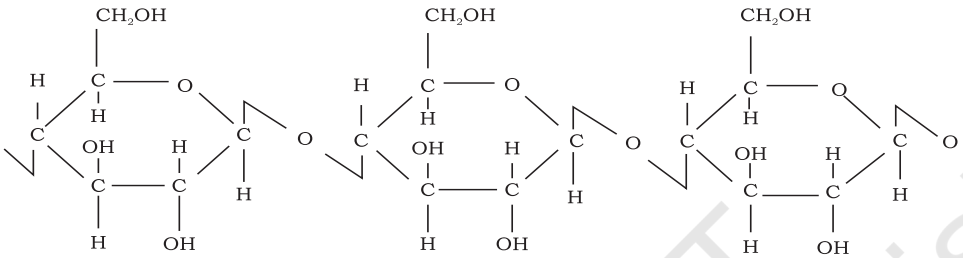
In prokaryotic cells, especially in bacteria, the cell wall is further covered with a heavily glycosylated protein known as **glycocalyx**. It acts as a barrier to invading pathogens and protects the cell from mechanical and ionic stresses. Glycocalyx is also involved in cell-cell interactions. In some cases, it could be present as a loose sheath called **slime layer** and in others, it could be thick and tough called **capsule**.

Among eukaryotes, the cell wall is mainly composed of polysaccharide (Fig. 2.5), which may be of **cellulose** (a linear polymer of glucose residues) for example, most higher plants, or **chitin** (a linear polymer of N-acetylglucosamine) for example, fungi. In plants, a growing cell is surrounded by a comparatively thin **primary cell wall**, having a scope for cell expansion. As it ceases growing, a new layer called **secondary cell wall** is formed between the primary cell wall and plasma membrane. A secondary cell wall is very rigid and thick compared to primary cell wall due to the deposition of **lignin**. A layer of calcium pectate (known as middle lamella) holds together neighbouring cells and connect their cytoplasm through a structure called **plasmodesmata**.

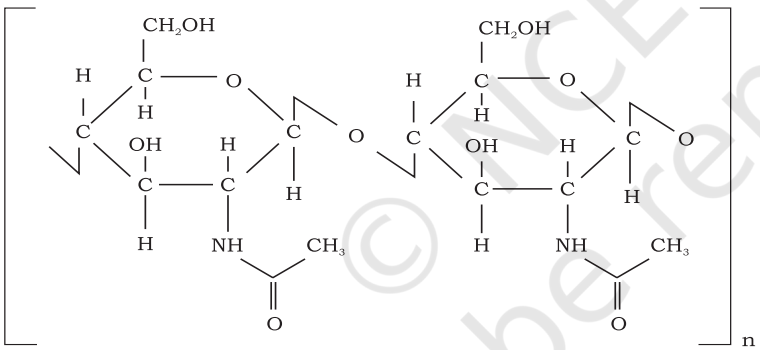
© NCERT not to be republished

2.3 ENDOMEMBRANE SYSTEM

Among eukaryotes, there are many cell organelles which are bound by membrane similar to that of cell membrane, and these are distinct in terms of structure and function. Yet, some membrane bound organelles work together in a system known as endomembrane (endo-'within') system, because their functions are co-ordinated with each other. It comprises a group of membrane bound organelles that work together in protein and lipid synthesis; its processing,



(a) Cellulose; Polysaccharide made up of repeating units of β-D-glucose linked via β(1→4) glycosidic linkages



(b) Chitin; Polysaccharide made up of repeating units of N-acetylglucosamine linked via β(1→4) glycosidic linkages

Fig. 2.5: Components of eukaryotic cell wall; (a) plant and (b) fungus

packaging and transport to their respective locations inside a cell (Box 2). Endomembrane system includes endoplasmic reticulum, Golgi complex, lysosomes and vacuole.

2.3.1 Endoplasmic Reticulum

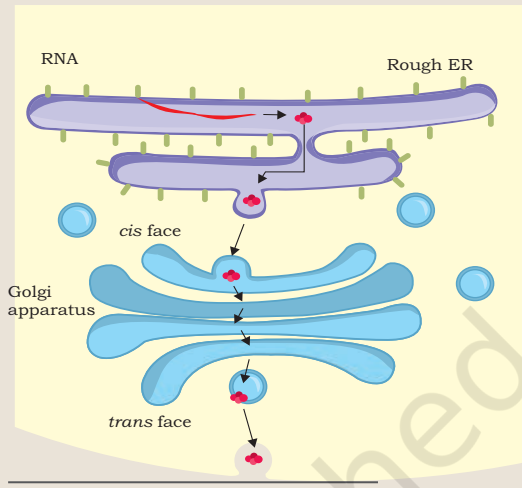
Endoplasmic reticulum (ER) is an extensive network of membrane-enclosed tubules and cisternae located near the nucleus and Golgi apparatus. It is exclusively present in a eukaryotic cell. ER is a large and dynamic structure

© NCERT not to be republished

Box 2

In 1960s, the pioneering work of George Palade and colleagues demonstrated the role of ER in protein processing and sorting. They revealed the entire pathway of secretory protein which is as follows:

Rough ER
↓
Golgi Apparatus
↓
Secretory vesicles
↓
Cell exterior



Pathway of secretory proteins

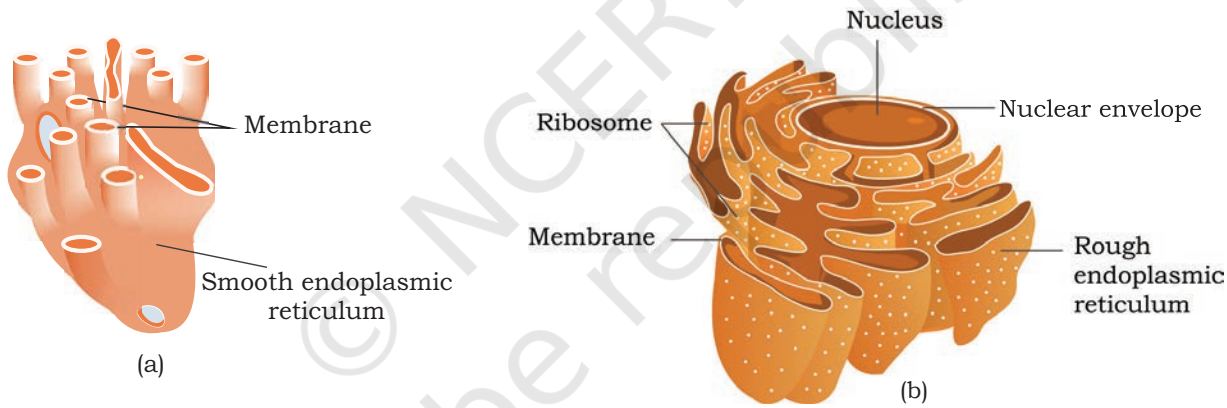


Fig. 2.6: Structure of endoplasmic reticulum (a) SER and (b) RER

continuously involved in protein synthesis, calcium storage and lipid metabolism. Based on the presence or absence of ribosomes, endoplasmic reticulum may either be **rough ER** or **smooth ER** (Fig. 2.6).

Rough Endoplasmic Reticulum (RER)

Rough ER can be distinguished from smooth ER by the presence of ribosomes on its cytosolic surface. Ribosomes are the **protein synthesising factory** of a cell (Refer to section 2.6). Proteins synthesised on free ribosomes are released into cytoplasm and they are directly transported

© NCERT not to be republished

to nucleus, mitochondria, chloroplast and peroxisomes to be used within the cell [Fig. 2.7(a)]. In case of bound ribosomes, after initiation of protein synthesis, the ribososme-protein complex is tranferred to a receptor on ER in eukaryotes. There the nascent protein under synthesis by ribosome is inserted into ER. These proteins may be either retained in the ER or transported to their destinations via Golgi complex through the secretory pathway [Fig 2.7(b)]. ER plays a significant role in trafficking of secretory proteins to Golgi apparatus, lysosomes and plasma membrane within the eukaryotic cells.

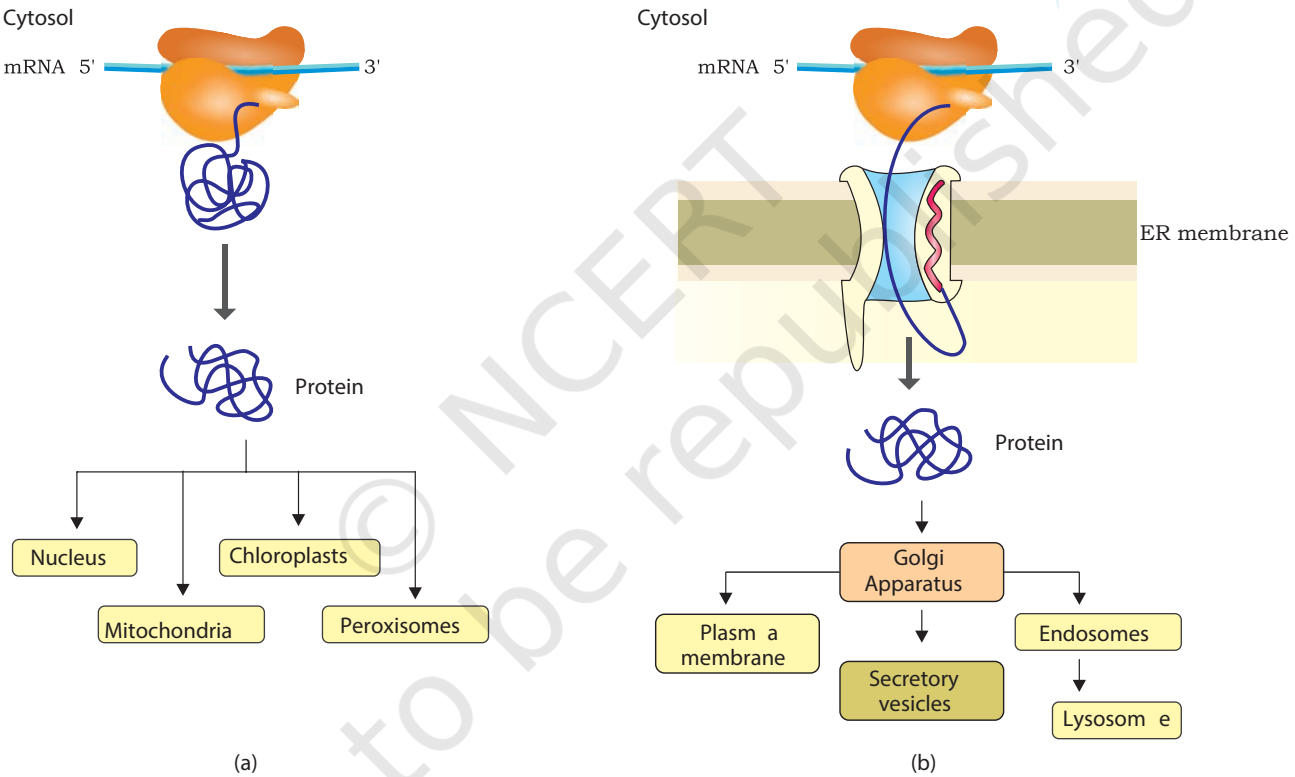


Fig. 2.7: Protein sorting in higher eukaryotes

Smooth Endoplasmic Reticulum (SER)

As mentioned earlier, ribosomes are not present on the surface of smooth ER [Fig. 2.6(a)]. It is mainly involved in lipid metabolism. As lipids are hydrophobic, they can not be synthesised in the cytosol. Most of the lipids are synthesised in the SER and transported to their respective destinations as transport vesicles. Phospholipids are one

© NCERT not to be republished

of the crucial components of a membrane, derived from glycerol. Its synthesis takes place on the outer side (cytosolic side) of SER membrane. SER is also an essential site for cholesterol synthesis.

2.3.2 Golgi Apparatus

First observed by Camillo Golgi, an Italian biologist in 1898, Golgi Apparatus (GA) is a dark stained reticular structure located near the cell nucleus. This was later found to be present in other cell types and named as Golgi apparatus or the Golgi complex. It is a membrane-bound cell organelle consisting of a series of flattened membranous sacs that look like stacked pouches called **cisternae**. Varied number of cisternae are present in a stack. The membrane of each cisterna in a stack separates its internal space from the cytosol. The Golgi cisternae are concentrically arranged near the nucleus with distinct *cis* face (cisternae nearest the endoplasmic reticulum) or the forming face and *trans* face (cisternae away from the endoplasmic reticulum) or the maturing face (Fig. 2.8). The Golgi apparatus principally performs the function of packaging the materials and prepare for secretions (receiving and shipping department). Transport vesicles move material from the ER to the Golgi complex.

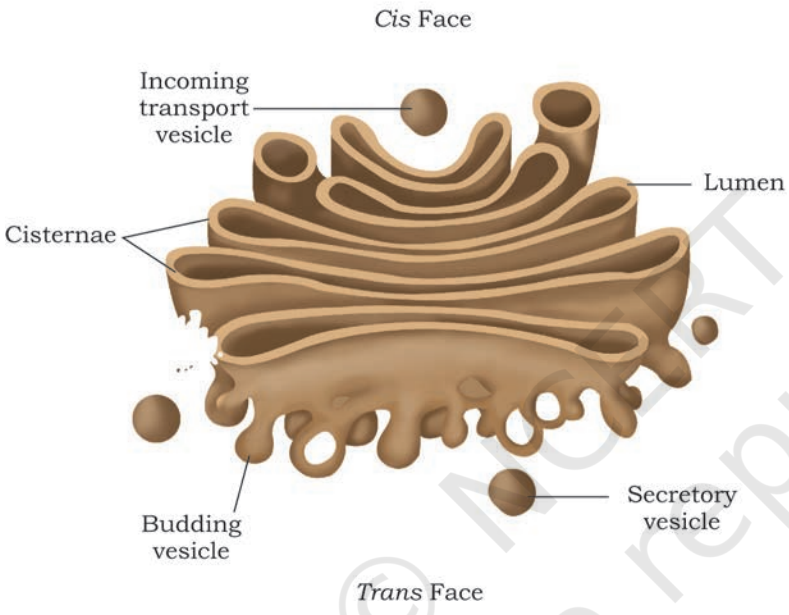


Fig. 2.8: Structure of Golgi apparatus

The material to be secreted moves from ER to the Golgi apparatus, during which vesicles are budded off from the ER. These vesicles travel to Golgi apparatus and fuse with *cis* face. The *trans* face gives rise to the vesicles that pinch off and travel to other sites (Box 2 at page 32). These vesicles can fuse with the plasma membrane, releasing their contents outside of the cell. Some vesicles deliver their contents to other organelles. Protein synthesised by ribosomes of the

© NCERT not to be republished

RER are modified in the cisternae of the Golgi apparatus before they are released from the *trans* face. For example, specific type of sugars are attached to some protein before they are released from the cell. Golgi apparatus is a central membrane organelle for trafficking and post-translational modification of protein and lipid in the cell. Also, it is an important site of formation for glycoproteins and glycolipids.

2.3.3 Lysosomes

Present in the cytoplasm, lysosomes are small spherical vesicles of approximately 0.2–0.5 micron in diameter, which are bound by a single membrane containing hydrolytic enzymes capable of breaking down macromolecules. These special vesicles of animal cells and some other eukaryotes are formed either from the Golgi apparatus or directly from the endoplasmic reticulum. Lysosomal enzymes show optimal activity at an acidic pH and thus are acid hydrolases, which are used in the dissolution and digestion of redundant structures or damaged macromolecules from within or outside the cell. For example, when an animal cell ingests food into a food vacuole, lysosomes fuse with the vacuole and break down the contents (carbohydrates, proteins, fats and other components) enzymatically. Lysosomes carry out intracellular digestion in a variety of circumstances. They also use their hydrolytic enzymes to recycle the cell's own organic material through a process called **autophagy** and the cell continually renews itself. In Tay-Sachs disease, brain becomes impaired due to accumulation of lipids in the cells because of lack of, or inactivation of lipid digesting enzymes in it.

2.3.4 Vacuoles

Vacuoles are membrane-bound intracellular organelle found in the cytoplasm of most plants, fungi and also some animal cells. The term 'vacuole' meaning 'empty' comes from its transparent morphology and lack of cytoplasmic material. Its major function is storage, structural support and recycling.

The number and size of vacuoles varies per plant cell. Young cells have large number of small vacuoles. As the plant cell matures, the vacuoles amalgamate to form a large central vacuole that occupies almost 90 per cent of the cytoplasmic volume. The central vacuole contains

© NCERT not to be republished

water, cell sap, solid inclusions and other metabolites. Other types of vacuoles found in a plant cell are lytic vacuoles, protein storage vacuoles (PSV) and other storage vacuoles. The vacuoles are covered by a membrane called the **tonoplast**.

The major function of vacuoles in plant cells is storage, maintaining cell turgor and also to protect cells during biotic stresses. Fungal vacuoles, however are comparatively complex organelle performing a variety of functions apart from storage such as helping in degradative processes, playing role in osmoregulation and maintains intracellular pH. The food vacuoles help in engulfing food particles and the contractile vacuoles in excretion. In case of amoeba, contractile vacuole plays a critical role in excretion and osmoregulation. In many protists, food vacuoles are formed by engulfing food particles.

2.4 MITOCHONDRIA

Mitochondria (singular: mitochondrion) are found in nearly all eukaryotic cells. Some cells may have single large mitochondrion, but more often a cell has hundreds or even thousands of mitochondria at variable location in cells depending upon the cell function. Mitochondria possess recognisable morphological characteristics despite considerable variability in appearance. A typical mitochondria is sausage shaped. In electron micrographs, they appear mostly as rod-shaped or cylindrical. They vary in size within the range 3.0 to 10.0 µm long and 0.5 – 1.5 µm wide.

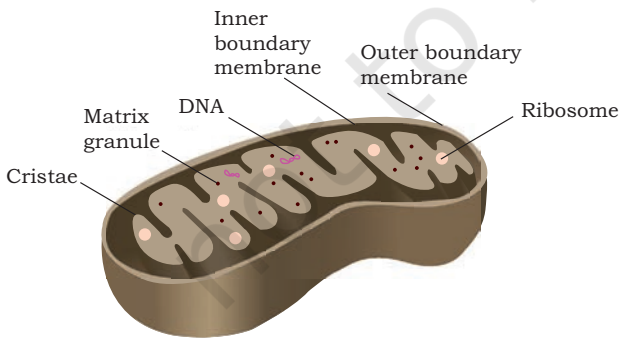


Fig. 2.9: Longitudinal section of mitochondria showing internal structure

Each mitochondrion is a double membrane bound structure i.e., outer and the inner membrane, each consisting of phospholipids bilayer with proteins (Fig. 2.9). The outer membrane is smooth, but the inner membrane has infoldings called **cristae** (singular: crista) that provides it comparatively larger surface area. The inner membrane divides the mitochondrion into two internal compartments. The narrow region between the inner and

© NCERT not to be republished

outer membrane is the **peri-mitochondrial space,** and the innermost compartment lined by inner membrane is called **mitochondrial matrix**. The inner membrane and matrix contains all enzymes and proteins involved in the process of tricarboxylic acid (TCA) cycle and cellular respiration for the purpose of ATP synthesis (details given in Chapter 5). Mitochondria also contains DNA molecules, ribosome (70S) and few RNA molecules. Some of the mitochondrial proteins are synthesised by genes present on the mitochondrial DNA. Therefore, mitochondrial matrix is the site of organellar DNA replication, transcription, protein synthesis and other enzymatic processes. We know that mitochondria consist of well over 1000 proteins which are varying within and between species in response to the needs of the organisms.

2.5 PLASTIDS

Plastids are usually present in the cytoplasm of plant cells. The term plastid is derived from the Greek word *Plastikas* (meaning formed or moulded). They are easily observed under the microscope as they are large. They bear some specific pigments, thus imparting specific colours to the plants. Three different types of plastids are recognised on the basis of their pigments or colouration.

a) **Chromoplasts**—These are coloured plastids containing variously coloured pigments such as yellow, red, pink, violet colours of flowers, fruits, leaves, etc. In chromoplasts, fat soluble pigments like carotene, xanthophylls and others are present.

b) **Leucoplasts**—These are colourless plastids. They are usually involved in the storage of various kinds of reserve food materials and are named accordingly, as amyloplasts (storage of starch), aleuroplasts (storage of protein) and elaioplasts or lipoplast (storage of oil).

c) **Chloroplasts**—These are the green plastids, universally

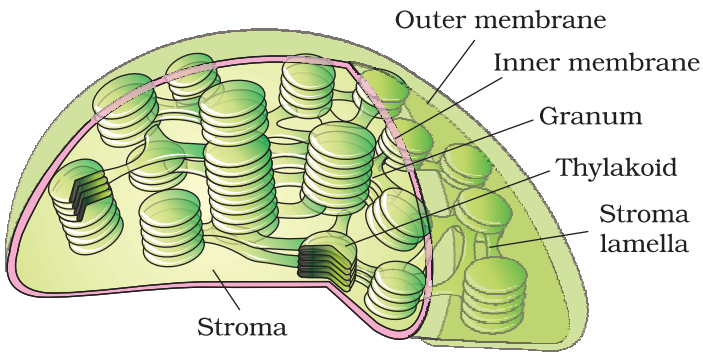


Fig. 2.10: Sectional view of chloroplast

© NCERT not to be republished

found in all the green parts of the plant, specially the green leaves. It contains large quantities of green pigments, called chlorophyll. Chlorophyll is a collection of four pigments, chlorophyll *a*, chlorophyll *b* and the yellow pigments—carotenoids and xanthophylls.

Chloroplasts are located predominantly in the mesophyll cells of the leaves. The chloroplasts of higher plants are generally lens-shaped such as oval, spherical, discoid, flat ellipsoids or even ribbon like organelles and in average 2 to 4 µm wide and 5 to 10 µm long. They are the largest organelles in the plant cell.

The chloroplast is bounded by a double membrane, separated by a very narrow intermediate space. Chloroplast also contains a third inner membrane which is organised into flattened membranous sacs, called **thylakoids**. Thylakoids are arranged in orderly stacks called **grana** (singular: granum), which resemble stacks of coins. Space (fluid) outside the thylakoids and within the chloroplast envelop is the **stroma**, which contains the chloroplast DNA and ribosomes as well as many enzymes. In addition, there are flat membranous tubules called the **stroma lamellae** connecting the thylakoids of the different grana. The membrane of the thylakoids enclose a space called a lumen. The membrane of the chloroplast divides the chloroplast space into three compartments: the intermediate space, the stroma, and the thylakoid space (Fig. 2.10).

Chlorophyll pigments are present in the thylakoids. The thylakoid membrane contains light-harvesting proteins, reaction centres, electron-transport chains and ATP synthase, which are the primary events of photosynthesis (detail of photosynthesis is given in Chapter 5). The ribosomes of the chloroplasts are smaller (70S) than the cytoplasmic ribosomes (80S).

2.6 RIBOSOMES

Ribosomes are '**protein synthesising factories**' scattered throughout the cytoplasm in both prokaryotic and eukaryotic cells. Each ribosome is a membraneless cell organelle, which was first observed by George Palade in 1955 under electron microscope.

© NCERT not to be republished

The number of ribosomes per cell are quite variable. A rapidly growing mammalian cell has approximately 10 million ribosomes. Ribosomes have been classified on the basis of their sedimentation rate in a centrifuge machine i.e., prokaryotic ribosome as 70S and eukaryotic ribosomes as 80S. It is peculiar to note that 70S ribosomes are also present in mitochondria and chloroplast of a eukaryotic cell, suggesting their relatedness to prokaryotic cell. Ribosomes are mainly composed of rRNA and proteins. Each ribosome has a small and a large subunit (Fig. 2.11). The 70S prokaryotic ribosome has larger 50S and smaller 30S subunits. On the other hand, 80S eukaryotic ribosomes also contain two subunits i.e., larger subunit 60S and smaller subunit 40S. In both prokaryotes and eukaryotes, the subunits of ribosomes remain dissociated in the cytoplasm when it is not involved in protein synthesis.

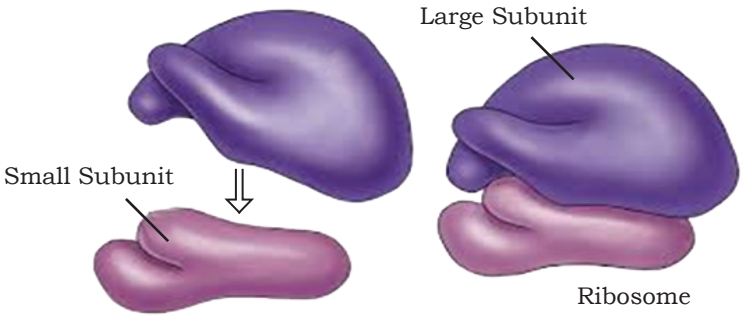


Fig. 2.11: Subunits of ribosome

In a ribosomal subunit, rRNAs assume characteristic secondary structure by utilising complementary base pairing and form a distinct three-dimensional structure by associating with ribosomal proteins. Some rRNAs present in the large subunit have catalytic activities and are called **ribozymes**.

Box 3

In ribosomes 70S and 80S, S denotes Svedberg unit, a unit for sedimentation rate. It is named after scientist Theodor Svedberg, a Swedish chemist who won Nobel Prize in chemistry for the invention of ultracentrifuge. Sedimentation rate is the measure of speed with which a particle sediments under gravitational field induced by a centrifuge.

2.7 MICROBODIES

These are small, single-membrane bound cell organelles, present only in eukaryotic cells. They are usually present near endoplasmic reticulum. Based on their functional properties, microbodies are classified into two types; peroxisomes and glyoxysomes.

© NCERT not to be republished

2.7.1 Peroxisomes

Peroxisomes are small, membrane-bound organelles. They are involved in energy metabolism in the cell, thus, serve as a site for enzymes involved in metabolic reactions. They are derived from ER and replicate by fission. Unlike mitochondria and chloroplast, peroxisome lacks its own genome.

In animal cells, peroxisomes perform two major functions: oxidation and lipid biosynthesis. They contain oxidases that help in peroxide production, and catalase that neutralizes the harmful products of oxidation reaction. In liver cells, peroxisomes also help in alcohol detoxification. In plant cells, peroxisomes have two major roles to play, i.e., conversion of fatty acids into carbohydrates in seeds and photorespiration in leaves.

2.7.2 Glyoxysomes

Glyoxysomes are specialised peroxisomes found in fungi and other higher plant (especially in fat storage tissues in germinating seeds). When oil filled seeds germinate, then the number and activity of glyoxysomes also increase. Glyoxysomes contain all the enzymes necessary for the fatty acid oxidation, glyoxylate cycle and gluconeogenesis. The seedling uses these sugars synthesised from fats until it is mature enough to produce them by photosynthesis. The conversion of lipids into glucose requires coordinated function of glyoxysomes, mitochondria and plastids.

2.8 CYTOSKELETON

Cytoskeleton is a multi-component system made of fibrous proteins. It maintains cell organisation and shape. It is a vital component of cell that provides mechanical support and plays a very crucial role, especially, during cell division, cell movement and intracellular transport. Cytoskeleton is composed of three major filaments varying in their protein composition and diameter: (1) Microtubules (25 nm) made of tubulin protein, (2) Actin filament (6 nm) made of actin protein and (3) Intermediate filament (10 nm) made of combination of different subunits of protein.

1. **Microtubules—** It is composed of globular proteins called tubulin (a dimer having α and β subunits). Tubulin proteins undergo polymerisation to form a

© NCERT not to be republished

protofilament. Microtubules are hollow rod-like, may contain 10–15 protofilaments (Fig. 2.12). In addition to its other functions, it is also responsible for rhythmic movement of cilia and flagella.

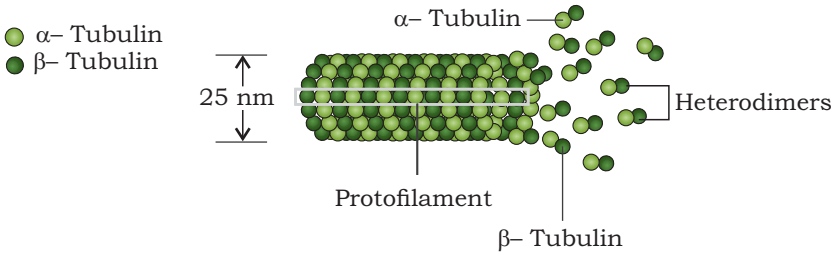


Fig. 2.12: Schematic representation of microtubules

2. **Actin Filament**— It is observed in skeletal muscle and plays a very important role in muscle contraction. It is richly found in the cytoplasm closer to the plasma membrane. Its main function is to provide strength to the cell, facilitating cytokinesis and cell movement.
3. **Intermediate Filaments**— These are strong filaments resembling a rope, primarily involved in providing mechanical strength to the cell.

2.9 CILIA AND FLAGELLA

Cilia (Singular: cilium) and flagella (Singular: flagellum) are hair-like, microscopic, filamentous protoplasmic structures. Both cilia and flagella are involved in cell motility. Although cilia and flagella are morphologically and physiologically identical, but it can be differentiated on the basis of their size, number and function (Table 2.1). Cilia are smaller in size and present in large number in a ciliated cell whereas, flagella are longer in size and typically vary from one or two in number. The body of a *Paramecium* (unicellular protozoan) is fully covered by a few thousands of cilia. Cells present in the upper respiratory tract of mammals bear cilia to expel particulate matter present in inhaled air. A mammalian spermatozoan contains single flagellum, while a unicellular *Chlamydomonas*, a green alga has two flagella.

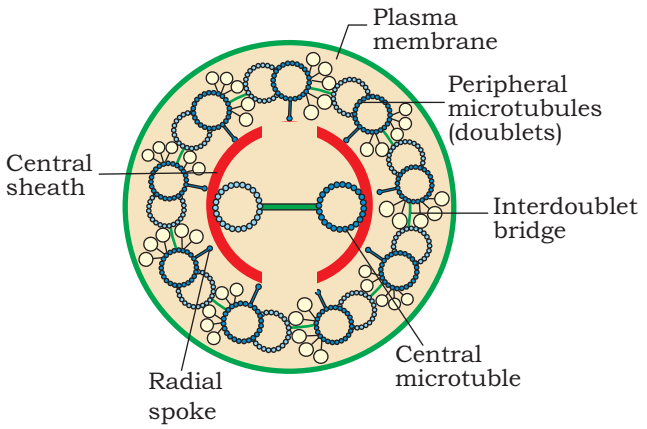


Fig. 2.13: Structure of eukaryotic cilium/flagellum

© NCERT not to be republished

Cilia and flagella are fibrillar and made up of microtubules. Their fundamental structure is same. Both arise from a centriole-like structure known as **basal body**. Electron microscopic view shows that they are bounded by a unit membrane (thickness 90 Å) which is continuous with plasma membrane (Fig. 2.13). They have a core known as **axoneme** in the matrix containing nine peripheral and two central microtubules. Such an arrangement is referred as **9+2 array**. Central fibrils are enclosed by a sheath. Prokaryotes also contain cilia/flagella but it differs structurally from a eukaryotic one in terms of 9+0 array.

Table 2.1: Difference between cilia and flagella

Characteristics	Cilia	Flagella
Size	Smaller in size up to 5–10 μm	Larger in size up to 150 μm
Location	Occurs throughout the surface of a cell	Occurs at one end of the cell
Number	Numerous in number	One or two in number
Movement	Moves in a co-ordinated rhythm and show sweeping or perpendicular stroke motion	Moves independently and show undulatory movement or whiplash movement
Examples	It is found in— • Protozoans (class—Ciliata) • Ciliated epithelium of metazoan • Larvae of Platyhelminthes, ribbon worms, Annelids, Mollusca and Echinodermata	It is found in— • Protozoans (class—Flagellata) • Sponges (Choanocyte cells) • Spermatozoa of Metazoa • Plants (algae and gamete cells)

2.10 CENTROSOME AND CENTRIOLES

Centrosome is present in animal cell cytoplasm near the nucleus. It consists of two cylindrical centrioles placed perpendicular to each other, embedded in amorphous pericentriolar materials. During cell division, it duplicates in S phase and separates to opposite direction during mitosis M-phase.

Centrioles are two cylindrical structures composed of nine triplets of microtubules of tubulin, arranged around a central cavity (Fig. 2.14). They act as a center of mitotic spindle assembly during cell division.

© NCERT not to be republished

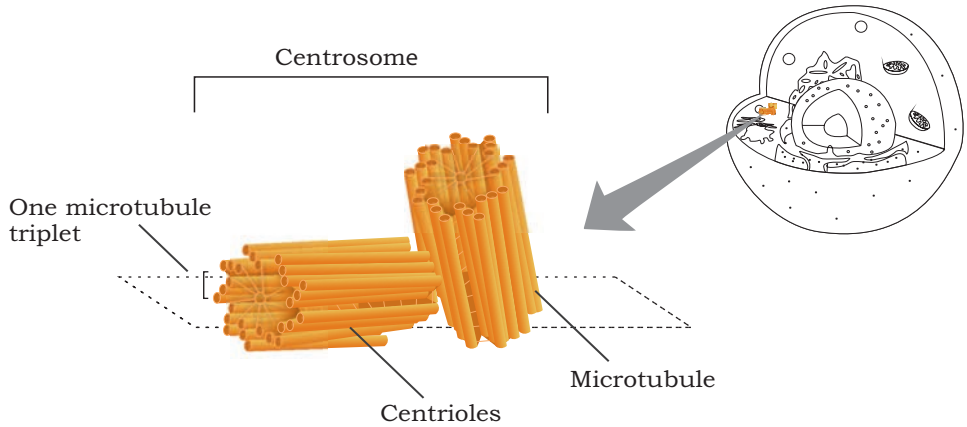


Fig. 2.14: Schematic view of centrioles

2.11 NUCLEUS

As compared to prokaryotes, eukaryotes have a well-defined nucleus (Fig. 2.15), a master controller of cell activities and a vast repository of genomic information. It contains genetic material and regulates gene expression by various mechanisms that are exclusive to eukaryotes.

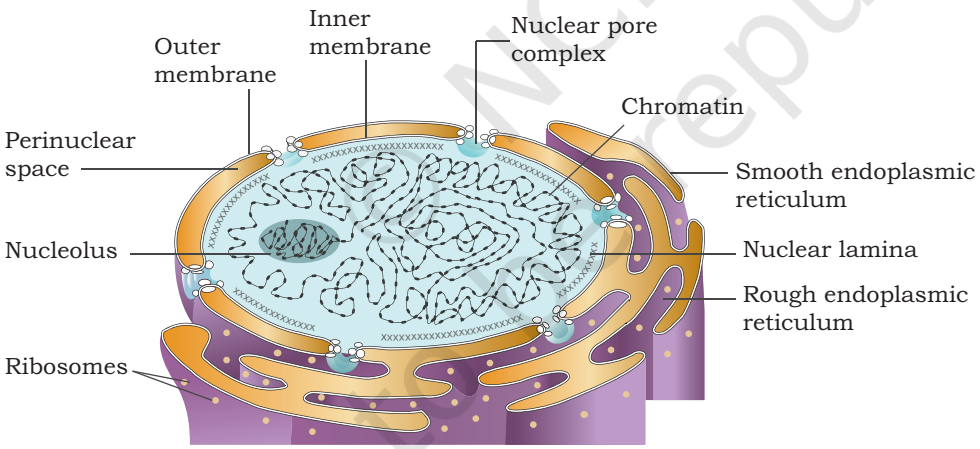


Fig. 2.15: Structure of a nucleus

2.11.1 The nuclear envelope

The nuclear envelope or nuclear membrane is a double lipid bilayer similar to the plasma membrane that surrounds the nucleus and acts as a barrier to keep molecules from freely moving between nucleus and cytoplasm. The nuclear envelope is porous, having pores called nuclear pores. At nuclear pores, the outer and inner nuclear membranes are continuous and each fitted with a proteinous nuclear

© NCERT not to be republished

pore complex. Each nuclear pore complex is composed of structural protein subunits called nucleoporins, arranged in a ring-like manner around a central channel (Fig 2.16). It allows trafficking of RNA and proteins. The outer nuclear membrane is continuous with the endoplasmic reticulum. A fibrous network of lamin proteins is present below the inner membrane called nuclear lamina that strengthens the structural framework of the nucleus.

Nuclear pore complexes are exclusively designed to transport proteins and RNA along with small polar and charged molecules.

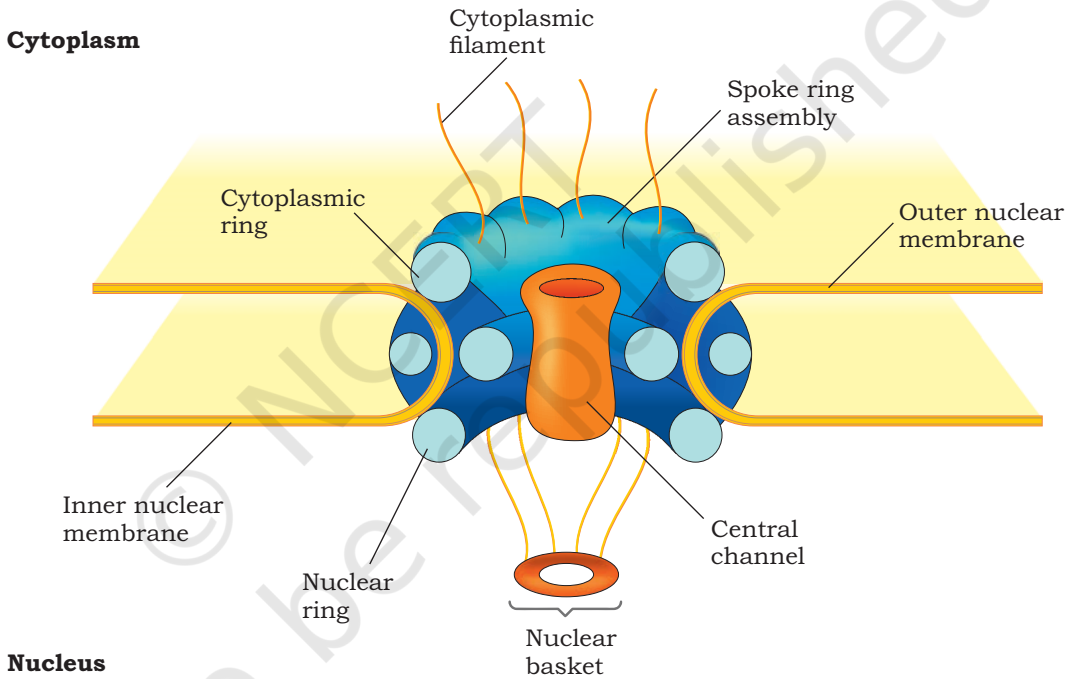


Fig. 2.16: Structure of a nuclear pore complex

2.11.2 Nucleoplasm

The nuclear envelope encloses a clear fluid called nucleoplasm or karyolymph or nuclear sap, which contains protein fibrils called nuclear matrix. It helps to maintain

© NCERT not to be republished

the shape of the nucleus. The enzymes associated with DNA replication and transcription are located in the nuclear matrix. Nucleolus and chromatin are suspended in the nucleoplasm.

2.11.3 Nucleolus

Nucleus comprises a distinct set of organelles referred to as **nuclear bodies**, but they lack a well-defined membrane. It helps in compartmentalising nuclear processes. Nucleolus (Fig. 2.15) is one of the most distinct nuclear bodies, involved in the synthesis of rRNA and ribosomes. Besides nucleolus, various other structures present are involved in many different activities that include transcriptional regulation, gene silencing, DNA repair, rRNA transcription and processing, and many others. The chromosomal region occupied by nucleolus comprises a large number of rRNA synthesising genes; therefore, it was named as **nucleolar organising region**.

2.12 CHROMOSOME

A chromosome is a thread-like microscopic structure formed by coiling of DNA packaged with protein containing all genetic material of an organism. Chromosomes can be categorised into two types: **autosomes** (body chromosome(s)) and **allosomes** (sex chromosome(s)). Certain hereditary traits are linked to a person's sex and are passed on through the sex chromosomes. Autosomes contain the rest of the genetic information. Cells in human being have 23 pairs of chromosomes (22 pairs of autosomes and one pair of sex chromosomes), with a total of 46 per cell. Table 2.2 shows the number of chromosomes in some plants and animals.

Box 4

Theophilus Painter, an American zoologist first declared the number of human chromosomes as 24 pairs or 48 in 1923 on the basis of microscopic studies which were wrong. It was corrected by Joe Hin Tjio, an Indonesia born American cytogeneticist in 1956 and declared the total number as 46. Every living organism including plants and animals has a fixed number of chromosomes.

© NCERT not to be republished

Table 2.2: Number of chromosomes in different eukaryotic organisms

Organism	No. of Chromosomes
Arabidopsis thaliana (diploid)	10
Maize (diploid)	20
Wheat (hexaploid)	42
Common fruit fly (diploid)	8
Earthworm (diploid)	36
Mouse (diploid)	40
Human (diploid)	46
Elephant (diploid)	56
Donkey (diploid)	62
Dog (diploid)	78
Gold Fish (diploid)	100
Tobacco (tetraploid)	18
Oat (hexaploid)	12

Prokaryotes like bacteria or blue-green algae usually contain single circular chromosome (called **nucleoid**) in the cytoplasm as they do not contain well-defined nucleus or other membrane-bound organelles. However, in some prokaryotes there may be more than one chromosome for example, *Vibrio cholerae.*

In eukaryotes, the chromosomes are present inside a well-defined nucleus. During the interphase of cell cycle, chromosomes are present in the form of long threads called **chromatin fibres**. The chromatin fibre is composed of nucleosomes. Nucleosomes consist of DNA wrapped around histone proteins. Chromatin enables the long strands of DNA molecules to fit into the cell nucleus. If all of the DNA molecules in a single human cell are unwound from the histones and placed end-to-end, they would stretch to 6 feet. During cell division, chromatin fibres condense further to form microscopically visible, long and slender chromosomes. The structure of chromosomes varies through the different phases of cell cycle. During cellular division, chromosomes are replicated (along

© NCERT not to be republished

with its DNA), divided, and passed successfully to their daughter cells. Sometimes, errors do occur leading to changes in the number or structure of chromosomes in new cells which may lead to serious problems.

SUMMARY

- Millions of tasks performed by our body become possible due to the presence of 'cell', which is the 'basic unit of life'. Cells can be broadly categorised into two types, prokaryotic (without membrane bound organelles and presence of nucleoid) and eukaryotic (with membrane bound organelles and presence of nucleus) cells.
- Both prokaryotic and eukaryotic cells are surrounded by a plasma membrane. Plasma membrane is mainly composed of phospholipids. It is selectively permeable and facilitates transport of molecules in and out of the cell. Besides plasma membrane, cells of bacteria, algae, fungi and some higher plants are surrounded by a rigid cell wall.
- Eukaryotic cell has two major compartments: nucleus and cytoplasm. Nucleus is enclosed by a nuclear envelope which has nuclear pores. The nuclear envelope encloses nucleoplasm, nucleolus and the genetic material in form of chromatin. Nucleolus helps in rRNA synthesis. A pair of centrioles form spindle apparatus during cell division in animal cells. Centrosome and centrioles form the basal body of the cilia and flagella which facilitate locomotion.
- Endomembrane system includes the endoplasmic reticulum (ER), Golgi apparatus, lysosomes and vacuoles. ER is constituted by tubules called cisternae. It is of two types: rough (with ribosomes) and smooth (without ribosomes). Ribosomes are non-membranous structures involved in protein synthesis.
- Ribosomes may be present freely in the cytoplasm or in bound state on rough ER.
- The function of the ER is to help in synthesis and transport of proteins, lipoproteins and glycogen.
- Golgi apparatus is a membranous organelle composed of flattened sacs. It performs the packaging of secretory substances and their transport from the cell.
- Lysosomes are single membrane structures containing enzymes for digestion of all types of macromolecules.

© NCERT not to be republished

- Vacuoles are membrane bound organelles that function in storage, structural support and recycling in the cell.
- Peroxisomes and microbodies take part in oxidation reactions within the cell. Glyoxysomes are peroxisomes involved in fat metabolism.
- Mitochondria are bound by two membranes. Its inner membrane has infoldings called cristae. The mitochondria help in oxidative phosphorylation and generation of ATP.
- In plant cells, pigment containing granules are called plastids. The plastids containing the green pigment, chlorophyll, are known as chloroplasts. Chloroplasts are essential for photosynthesis.

EXERCISES

1. The Fluid Mosaic Model has been proposed by
   (a) Robert Brown
   (b) Schleiden and Schwann
   (c) Robert Virchow
   (d) Singer and Nicolson
2. Ribosomes are composed of
   (a) only rRNA
   (b) rRNA and proteins
   (c) rRNA, proteins and DNA
   (d) lipids, proteins and DNA
3. Tonoplast is
   (a) a membrane covering the cell wall in plant cells
   (b) the inner membrane of the mitochondria
   (c) a membrane covering the vacuoles
   (d) a membrane covering the plastids
4. Describe the various mechanisms of transport across plasma membrane with the help of labelled diagrams.

© NCERT not to be republished

5. Match the following

Column I	Column II
(a) Nucleolus	(i) Alcohol detoxification
(b) Mesosome	(ii) Infoldings of inner mitochondrial membrane
(c) Vacuoles	(iii) Protein synthesis
(d) Cristae	(iv) Disc shaped sacs in Golgi
(e) Ribosomes	(v) rRNA synthesis
(f) Thylakoid	(vi) Membranous extensions of plasma membrane
(g) Peroxisomes	(vii) Storage and structural support
(h) Cisternae	(viii) Membranous sacs in chloroplast

6. What is the significance of the ratio of protein and lipids in membranes? How does varying the concentration of lipids in a membrane affect its function?
7. State the importance of cell wall in prokaryotic cells.
8. A eukaryotic cell contains organelles which may be bound by a single-membrane; double-membrane or non-membrane bound organelles. Classify the various eukaryotic organelles into these three types.
9. Mention the different types of vacuoles.
10. Peroxisomes share similarities as well as differences with mitochondria and chloroplast. Comment.
11. What are glyoxysomes? Where are these present? Mention their functions.
12. Cell is the structural and functional unit of life. Justify the statement.
13. Distinguish between
    (a) cilia and flagella
    (b) primary and secondary cell wall
    (c) lysosomes and vacuoles
    (d) microtubules and actin filaments
    (e) active and passive transport

© NCERT not to be republished

11150CH03

Chapter 3

Biomolecules

3.1 *Carbohydrates*
3.2 *Fatty Acids and Lipids*
3.3 *Amino Acids*
3.4 *Protein Structure*
3.5 *Nucleic Acids*

In the previous chapter you have learnt about the cell and its organelles. Each organelle has distinct structure and therefore performs different function. For example, cell membrane is made up of lipids and proteins, cell wall is made up of carbohydrates, chromosomes are predominantly made up of proteins and DNA, and ribosomes are made up of proteins and RNA. These ingredients of cellular organelles are also called **macromolecules** or **biomolecules**. There are four major types of biomolecules— carbohydrates, proteins, lipids, and nucleic acids. Apart from being structural entities of the cell, these biomolecules play important functions in cellular processes. In this chapter you will study the structure and functions of these biomolecules.

3.1 Carbohydrates

Carbohydrates are one of the most abundant classes of biomolecules in nature and found widely distributed in all life forms. Chemically, they are aldehyde and ketone derivatives of the polyhydric alcohols. Major role of carbohydrates in living organisms is to function as a primary source of energy. These molecules also serve as energy stores,

© NCERT not to be republished

metabolic intermediates, and one of the major components of bacterial and plant cell wall. Also, these are part of DNA and RNA, which you will study later in this chapter. The cell walls of bacteria and plants are made up of polymers of carbohydrates. Carbohydrates also act as informational materials and linked to surfaces of proteins and lipids to act in cell–cell interaction, and in the interaction between cells with other elements in the cellular environment where they play a role.

(A) Classification of carbohydrates

Carbohydrates are found in various forms ranging from simple sugars to complex polymers of more than one unit, and accordingly these have been classified. These are commonly classified into three categories namely, monosaccharides, oligosaccharides, and polysaccharides.

1. Monosaccharides

Monosaccharides are simple sugars which cannot be further hydrolysed into simpler forms. These monosaccharides contain free aldehyde (-CHO) and ketone (>C=O) groups, with two or more hydroxyl (-OH) groups with a general formula of $C_n(H_2O)_n$. Based on the number of carbon atoms and functional groups, monosaccharides are classified as given in Table 3.1.

Table 3.1: Classification of monosaccharides

S. No.	Class of monosaccharides based on number of carbon atoms	Class of monosaccharides based on functional groups	
		Aldoses	**Ketoses**
1.	Trioses ($C_3H_6O_3$)	Glyceraldehyde (an aldotriose)	Dihydroxyacetone (a ketotriose)
2.	Tetroses ($C_4H_8O_4$)	Erythrose	Erythrulose
3.	Pentoses ($C_5H_{10}O_5$)	Ribose	Ribulose
4.	Hexoses ($C_6H_{12}O_6$)	Glucose	Fructose

2. Oligosaccharides

Conventionally, oligosaccharides are carbohydrates having two to ten units of monosaccharides joined together by glycosidic bond. Some commonly occurring oligosaccharides are maltose, lactose, sucrose, etc., which are disaccharides (having two monosaccharides).

© NCERT not to be republished

3. Polysaccharides

Polysaccharides are polymers of more than ten monosaccharide units joined together by glycosidic linkages. These are classified in a number of ways depending upon the type of repeating monosaccharide unit (homo- and hetero-polysaccharides); in the degree of branching, and in the type of glycosidic linkage between the monomeric units. Examples of some common polysaccharides are starch, glycogen, cellulose, and chitin.

Carbohydrates can be conjugated to proteins and lipids to form **glycoconjugates**. There are three types of glycoconjugates; **glycoproteins, proteoglycans** and **glycolipids**. If the protein component is predominant in the association of carbohydrate and protein, it is called glycoprotein. If the association contains major amount of carbohydrate than protein, then it is called proteoglycan. When the carbohydrate conjugates with lipids, it is called glycolipid.

(B) Structure and properties of carbohydrates

(a) Monosaccharides

Structure of some common monosaccharides are given in (Fig 3.1). Monosaccharide such as glucose exists both as

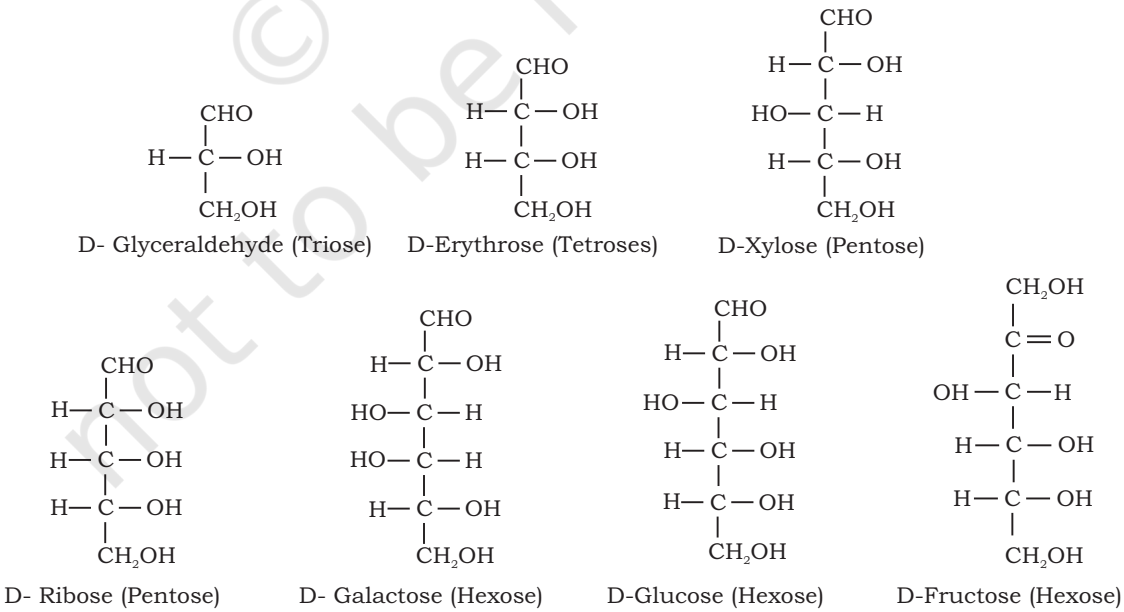


Fig. 3.1: Structure of some monosaccharides

© NCERT not to be republished

straight chain structure and cyclic structure (Fig. 3.2). Cyclic structures are the result of hemiacetal formation by intramolecular reaction between carbonyl group and a hydroxyl group.

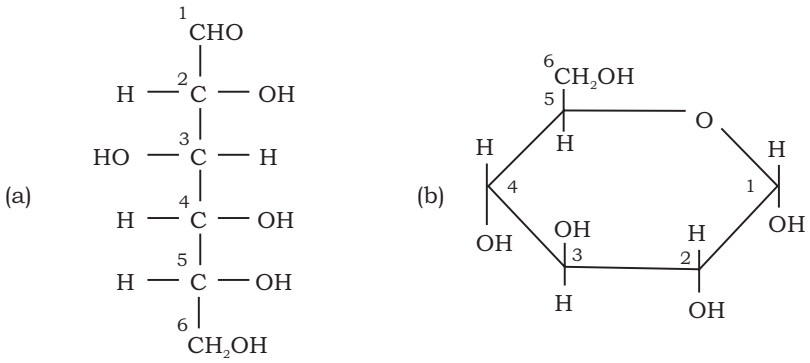


Fig. 3.2: Structure of glucose: (a) straight chain and (b) cyclic form

All monosaccharides except dihydroxy acetone contain one or more asymmetric (chiral) carbon (carbon atoms bound by four different groups), thus, are optically active isomers **(enantiomers)**. A molecule with *n* chiral centres can have 2^n stereoisomers. Thus, glyceraldehyde with one chiral centre has 2^1=2 and glucose with four chiral centres has 2^4=16 stereoisomers.

The orientation of the –OH group that is most distant from the carbonyl carbon determines whether the sugar belongs to D or L sugars. When this –OH group is on the right side of the carbon atom bearing it then the sugar is D-isomer, and when it is on the left, the sugar is L isomers (Fig. 3.3). Most of the sugars present in the biological system are D sugars.

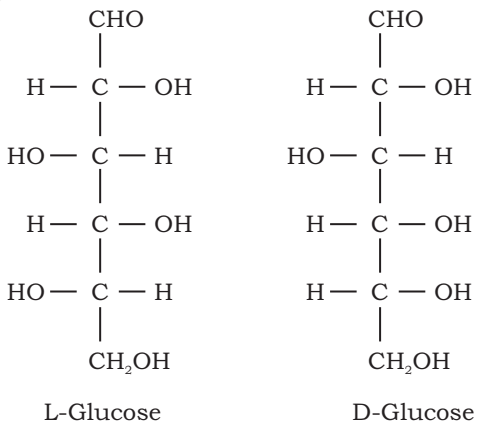


Fig. 3.3: L and D forms of glucose

Isomeric forms of monosaccharide that differ only in their configuration about the hemiacetal (formed due to reaction between alcoholic and aldehyde groups of a monosaccharide) or hemiketal (formed due to reaction between alcoholic and keto groups of a monosaccharide) carbon atom are called **anomers**. The carbonyl carbon atom is called the anomeric carbon. In the α - anomer, the -OH group of the anomeric carbon is on the opposite of the sugar ring from CH_2OH group at the chiral centre that designates the D and L configuration (C-5 in case of glucose). The other anomer is

© NCERT not to be republished

known as β-anomer. The interconversion of α and β anomers in aqueous solution is called **mutarotation**, in which one ring form opens briefly into the linear form, then closes again to produce β anomers (Fig 3.4).

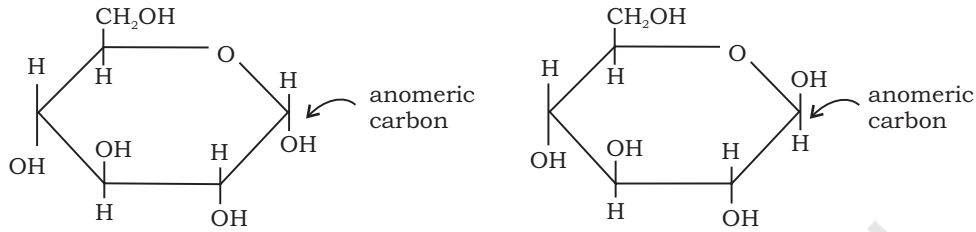


Fig. 3.4: Two cyclic forms of glucose

Isomers having different configuration of –OH only at one carbon atom are known as epimers. The most important **epimers** of glucose are mannose (epimers at C-2) and galactose (epimers at C-4) as shown in Fig. 3.5.

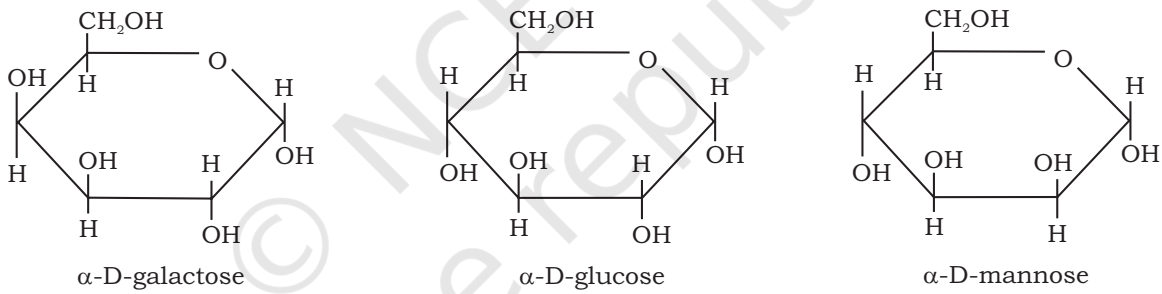


Fig. 3.5: The epimers of glucose

(b) Oligosaccharides

Oligosaccharides consist of two to ten monosaccharides joined by glycosidic linkage. For example, disaccharide maltose contains two D-glucose residues joined by a glycosidic linkage, which is a covalent bond formed by joining of –OH group of one monosaccharide with the anomeric carbon of the other sugar unit. Lactose is made up of D-galactose and D-glucose residues (Fig. 3.6 and 3.7).

Disaccharides can be hydrolysed to yield their constituent monosaccharides by boiling with dilute acid. For example, hydrolysis of sucrose yields a mixture of glucose and fructose. Other oligosaccharides are

© NCERT not to be republished

trisaccharides (for example, raffinose, which is made up of glucose, galactose, and fructose units), tetra-, penta- and hexasaccharides.

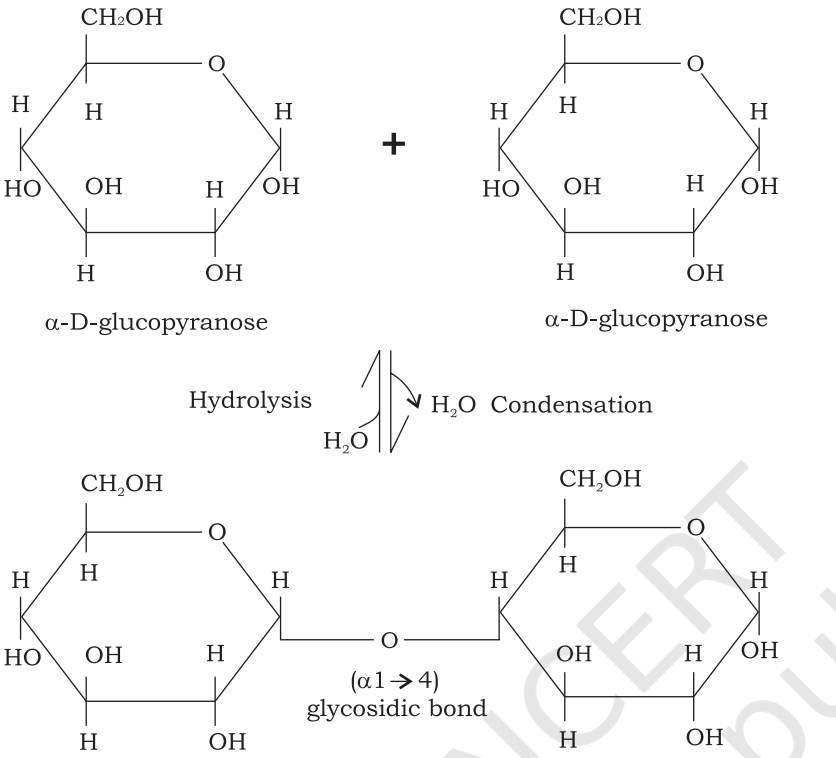


Fig. 3.6: Formation of maltose from two molecules of α D-glucose

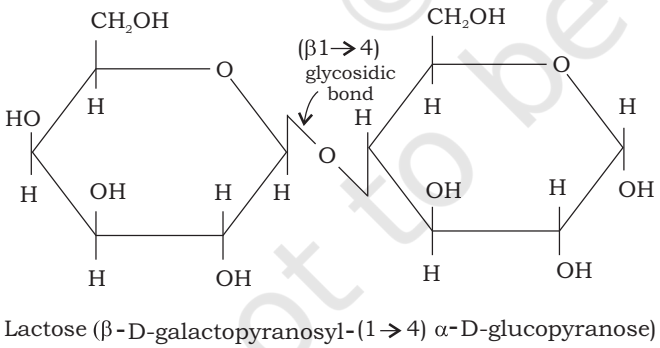


Fig. 3.7: Structure of lactose

(c) Polysaccharides

Depending on the functional roles played by homopolysaccharides are divided into storage polysaccharides and structural polysaccharides. Storage

© NCERT not to be republished

polysaccharides serve as storage form of monosaccharide that is used as fuel. Starch is an example of storage polysaccharide in plants, and glycogen is the storage polysaccharide in animals. Structural polysaccharides such as cellulose and chitin serve as structural elements in plant cell wall and animal exoskeleton, respectively. Heteropolysaccharides, unlike homopolysaccharides, provide extracellular support for organisms. In extracellular space of animal tissues, these form a matrix that holds individual cells together and provides shape, support and protection to the cells and tissues. The names and properties of some homopolysaccharides are given in Table 3.2

Table 3.2: List of some common homopolysaccharides

Name	Constituent monosaccharide	Biological significance
Starch	α-D-glucose	Storage of energy in plants
Glycogen	α-D-glucose	Storage of energy in bacteria and animals
Cellulose	β-D-glucose	It plays structural role and provides rigidity and strength to the cell wall.
Chitin	β-N-acetyl-D-glucosamine	It plays structural role and provides rigidity to exoskeleton of insects.
Inulin	β-D-fructose	Storage of energy in plants.
Pectin	α-D-galacturonic acid	It has structural role: holds cellulose fibrils together in plant cell walls.
Dextran	α-D-glucose	It plays structural role as an extracellular adhesive in bacteria.
Xylan	β-D-xylose	It has storage and supporting roles in plants.

Examples of some common polysaccharides

(a) Starch

Starch occurs in plants as reserve carbohydrate in tubers, seeds, fruits, and roots. It is composed of two homopolysaccharides, amylose (15-20%) and amylopectin

© NCERT not to be republished

(80–85%). Amylose is a linear polymer of α-D-glucose monomers and linked by α(1→ 4) bonds (Fig. 3.8). Amylopectin, on the other hand, consists of glucose units linked by α(1→ 4) glycosidic linkage, like amylose. However, unlike amylose, it is highly branched. Branch points occur every 24 to 30 glucose residues and linkage at the branch points is α(1→ 6) glycosidic (Fig. 3.9). Characteristic blue colour of the starch with iodine is due to amylose. In contrast, amylopectin gives only dull reddish brown colour with iodine. The enzymes present in saliva (salivary amylase) and pancreatic juice (pancreatic amylase) hydrolyses α(1→ 4) glycosidic linkages of starch therefore digesting it into monomeric glucose residues.

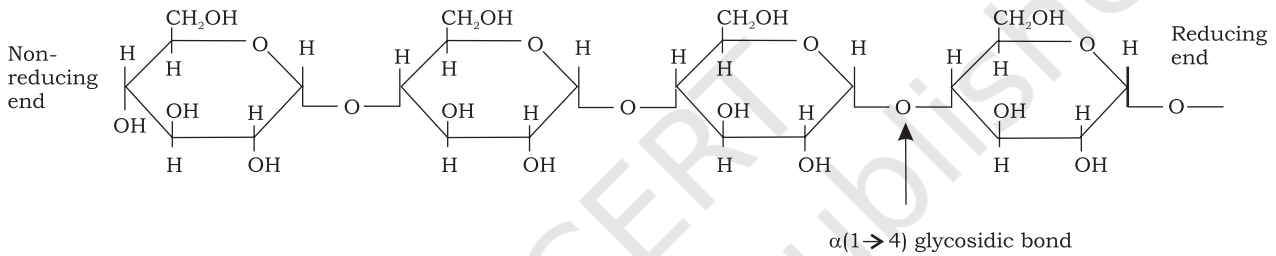


Fig. 3.8: Structure of amylose

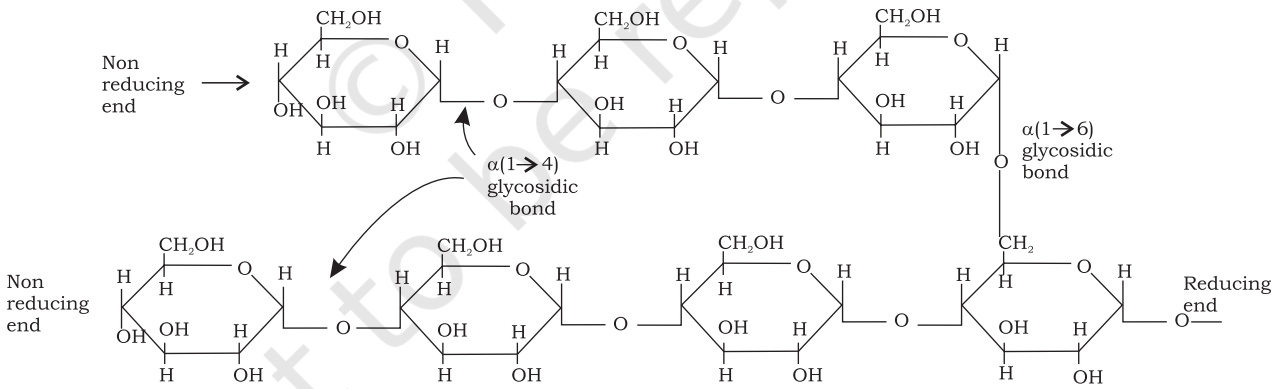


Fig. 3.9: Structure of amylopectin

(b) Glycogen

Glycogen is an extensively branched storage homopolysaccharide found in animals. Similar to amylopectin, it consists of glucose units joined together by α(1→ 4) glycosidic linkage and having α(1→ 6) linkage at branching points. Muscle cells contain glycogen at

© NCERT not to be republished

1–2 per cent of their dry weight, and liver cells contain up to 10 per cent of their dry weight as glycogen.

(c) Cellulose

Cellulose is the most abundant extracellular structural polysaccharide found in plants. It is also the most abundant of all biomolecules in the biosphere. It is the primary structural component of plant cell wall. Structurally, cellulose is a linear polymer of upto 15000 D-glucose units linked by β(1→ 4) glycosidic bonds (Fig. 3.10). Unlike starch, cellulose cannot be digested by humans as human gut lacks the β(1→ 4) glycosidic bond hydrolysing enzyme known as cellulase. However, cattles and termites can digest cellulose as their gut harbors symbiotic microorganisms that secrete cellulase which hydrolyses and digests cellulose.

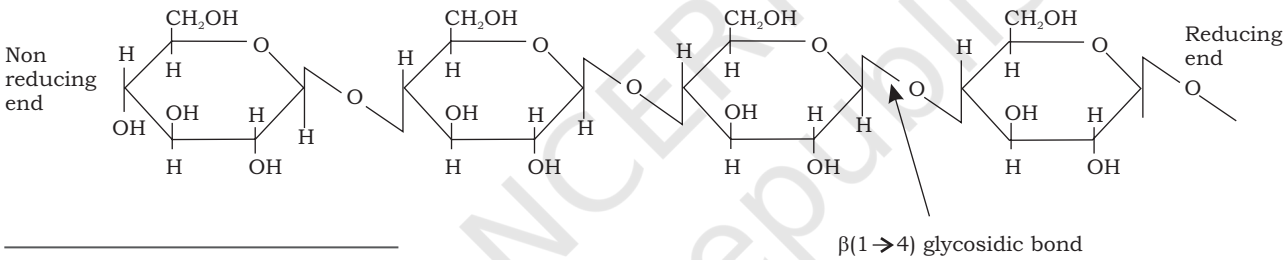


Fig. 3.10: Structure of cellulose

(d) Chitin

Chitin is a linear polysaccharide of β(1→ 4) linked N-acetyl-D-glucosamine residues. It is the main structural component of the exoskeleton of invertebrates (crustaceans, insects and spiders), and main component of cell walls of most fungi. Chitin and cellulose have similar structures except that OH group at second carbon position of cellulose is replaced by an acetamido group in chitin. Extensive hydrogen bonding of N-acetyl side chains makes chitin tough and insoluble polymer.

(e) Peptidoglycan

Peptidoglycan constitutes the rigid component of bacterial cell wall. It is heteropolysaccharide of alternating β(1→ 4) linked N-acetyl muramic acid (NAM) and N-acetyl-D-glucosamine (NAG) residues. The linear polysaccharide chains are cross linked by short peptides attached to N-acetyl muramic acid. Cross linking by peptide welds the

© NCERT not to be republished

polysaccharide chains into a strong sheath that envelops the entire cell and prevents osmotic rupture of the cell. Lysozyme, which is an enzyme present in human tears kills bacteria by hydrolysing the β(1→ 4) glycosidic linkage of peptidoglycan.

Box 1

Agar

Agar is gelatinous heteropolysaccharide present in the cell wall of marine red algae such as species of *Gelidium*, *Gracilaria*, *Gigartina,* etc. It is a mixture of sulphated heterpolysaccharides made up of D-galactose and L-galactose derivatives ether-linked between C3 and C6. Agarose is the agar component with very few charged groups (sulfates, pyruvates). It has molecular weight in the range of 80,000-1,40,000. If agar and agarose are dissolved in hot water, they form solution which upon cooling sets to a gel. Agarose gels are used as inert support for the electrophoretic separation of nucleic acids. Agar is used to form a surface for the growth of bacterial and plant tissue cultures.

3.2 Fatty Acids and Lipids

Lipids are a group of organic compounds found in living organisms. They vary in their structures and functions. Because of their hydrophobic and non-polar nature, lipids are soluble in organic solvents. Lipids are primarily made up of hydrocarbon chains connected to glycerol *via* ester linkage. We broadly classify lipids into two categories– simple lipids and compound lipids. Various types of lipids are included within these two major categories of lipids. These include fats, triacylglycerols, wax, phospholipids, steroids, etc.

Fatty acids are obtained as a result of hydrolysis of fats. Naturally occurring fatty acids are generally synthesised from two carbon units and hence, contain even number of carbon atoms (Fig. 3.11). Synthesised from 2 carbon units, fatty acid chains may be saturated (having no double bonds) or unsaturated (having one or more double bonds) (Table 3.3).

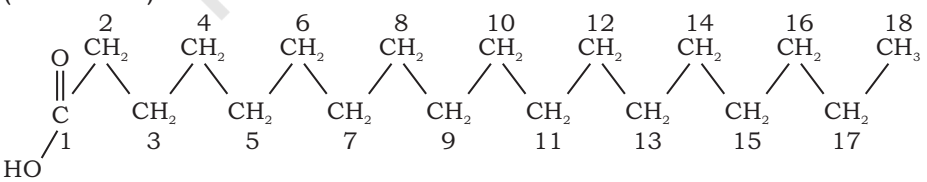


Fig. 3.11: Structure of a fatty acid (stearic acid)

© NCERT not to be republished

The numbering of C-atoms in a fatty acid hydrocarbon chain starts from carboxylic carbon and ends at methyl carbon in increasing order. For example in Fig. 3.11 and Fig. 3.12, carboxyl carbon is designated as 1 and methyl carbon is numbered 18. Fatty acids are denoted by total number of carbons followed by colon (:) and then total number of double bonds with Δ (delta) having superscript number defining the position of the double bond in parenthesis. For example, a fatty acid with 18 carbon with two double bonds between C-9 and C-10: and C-12 and C-13 will be denoted as 18:2 ($\Delta^{9,12}$) (Fig. 3.12).

Unsaturated fatty acids are of two types based on the degree of unsaturation as follows:

Monounsaturated fatty acids: 'Mono' means single. Therefore, monounsaturated fatty acid contains only a single double bond. For example, oleic acid.

Polyunsaturated fatty acids: As the name suggests, these fatty acids contain more than one double bond. e.g., Linoleic acid contains two double bonds (Fig. 3.12); Linolenic acid contains three double bonds; and Arachidonic acid contains four double bonds.

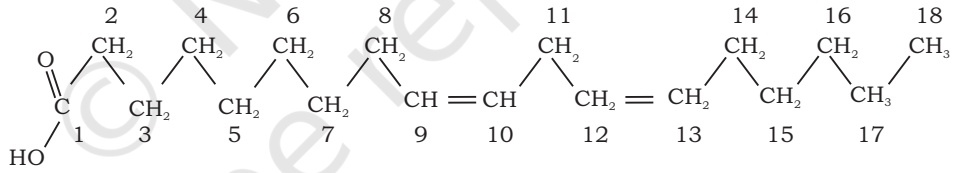


Fig. 3. 12: Structure of polyunsaturated fatty acid (linoleic acid) [18:2 (Δ9,12)]

Table 3.3: Some examples of saturated and unsaturated fatty acids indicating their chain length and melting point

Common name	Symbol	No. of Carbons
Saturated		
Lauric acid	12:0	12
Myristic acid	14:0	14
Palmitic acid	16:0	16
Stearic acid	18:0	18
Arachidic acid	20:0	20
Unsaturated		
Palmitoleic acid	16:1 (Δ^9)	16

© NCERT not to be republished

Oleic acid	18:1 (Δ^{9})	18
Linoleic acid	18:2 ($\Delta^{9, 12}$)	18
Arachidonic acid	20:4 ($\Delta^{5, 8, 11, 14}$)	20

Classification of lipids

1. Simple Lipids

A simple lipid is a fatty acid ester having various alcohols with no other group. For example, fats and waxes.

(a) Triacylglycerols: Widely known as triglycerides (or neutral fats), these compounds are the esters of glycerol and fatty acids (Fig. 3.13). The triglycerides containing the same fatty acids in all three ester positions are called simple triglycerides. The triglycerides contain more than one type of fatty acid at its three positions called mixed triglycerides. Triacylglycerol is essential for providing energy to body. Triglycerides are a vehicle of energy storage, primarily in adipose tissue.

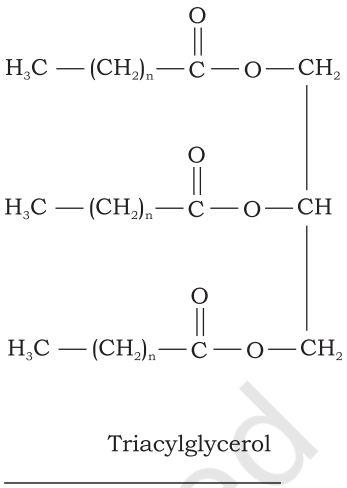


Fig. 3.13: Structure of triacylglycerol

(b) Waxes: Wax is formed as a result of esterification of fatty acids with a monohydric alcohol of high molecular weight. Waxes are found in a wide variety of organisms. For example, wax serves as a protective coating on the surface of animals and plants, and also reduces water loss in tropical plants.

2. Compound lipids

Compound lipids are fatty acid esters with alcohols and contain additional groups. The fatty acid components of these lipids make hydrophobic tails, and the alcohol moiety along with additional group constitutes hydrophilic head. Such a structure is known as amphipathic molecules. Common examples of compound lipids include phospholipids and steroids.

Phospholipids are amphipathic molecules made up of two hydrophobic fatty acid tails and a hydrophilic phosphate group head. Phospholipids are primarily found in cell membranes. These compounds are composed of fatty acid chains attached to a glycerol backbone **(glycerophospholipids)** or sphingoid base backbone **(sphingophospholipid)** (Fig. 3.14). A modified phosphate group occupies the third carbon of glycerol in glycerophospholipids. A phospholipid is defined by the

© NCERT not to be republished

kind of modifier attached to the phosphate group. The most common are choline (phosphatidylcholine) and serine (phosphatidylserine). Similarly, in the case of sphingolipids, when the head group contains only a hydrogen atom, it is called ceramide. Many a time, the head group in sphingolipids can be phosphocholine, yielding a sphingomyelin. In sphingolipids, the fatty acid chain is attached to the sphingoid backbone via amide linkage instead of ester linkage.

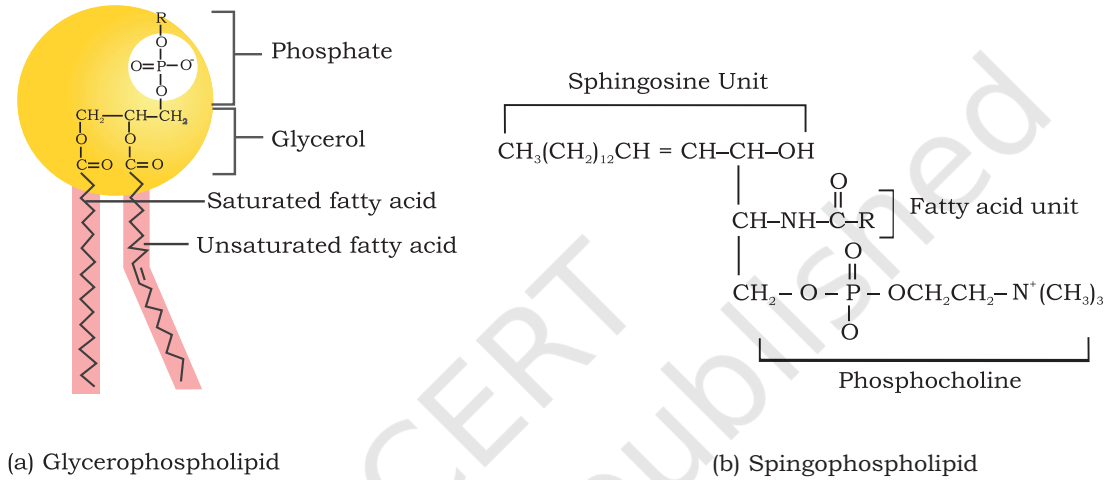


Fig. 3.14: Structure of (a) glycerophospholipid and (b) sphingophospholipid

Phospholipids do not spontaneously mix with water. Instead, they acquire a sphere-shaped structure known as **micelle**.

Steroids are distinct from other lipids in having a peculiar four-fused ring structure. However, like all the other types of lipids discussed so far, steroids are hydrophobic in nature and insoluble in water. In a cell, steroids act as receptor ligands and help to control the metabolism. **Cholesterol** is the most common derivative of steroids (Fig. 3.15). Synthesised primarily by the liver, cholesterol is the key precursor of all the steroid hormones, including testosterone and estradiol. Besides, cholesterol is also found in the plasma membrane of most eukaryotes, where it provides rigidity. Common examples of steroids found in plants are phytosterols

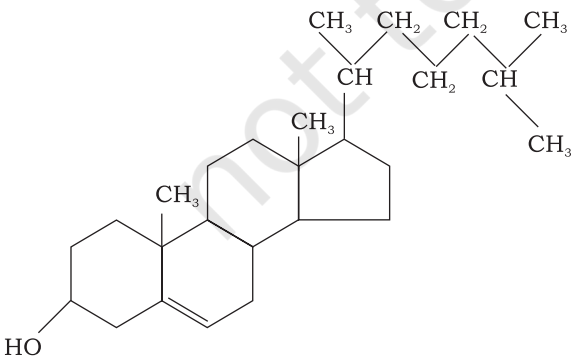


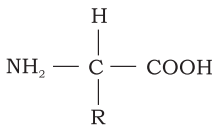
Fig. 3.15: Structure of cholesterol

© NCERT not to be republished

and stigmasterols which regulate membrane fluidity and permeability. Ergosterol is an important precursor of vitamin D and is typically found in fungi.

3.3 AMINO ACIDS

Amino acids, the building blocks of proteins, are represented by general formula:



where central carbon atom is called α-carbon, which is linked to four different groups; an acidic carboxylic (-COOH), a basic amino (-NH_2) group, a hydrogen atom, and an R group called side chain. Only the side chain R varies in all 20 amino acids, it can be as simple as a hydrogen atom (H) in glycine or it can be a methyl (-CH_3) in alanine. The carboxylic group contributes the first carbon, and the carbon atom to which carboxylic group is attached is called α-carbon. Since the α-carbon of most amino acids is tetrahedrally attached by four different groups, hence α-carbon of amino acids is chiral or asymmetric. Because of this asymmetric α-carbon, amino acids are present in two optically active forms (L and D) or mirror image forms (Fig. 3.16).

In L isomers –NH_2 group is present to the left and in D isomers –NH_2 group is present to the right of α-carbon. Only L isomers of amino acids are found in proteins, D isomers are rare in biological protein. List of the 20 standard amino acids is given in Table 3.4. Structure of the 20 amino acids are given in Fig. 3.17 to 3.21.

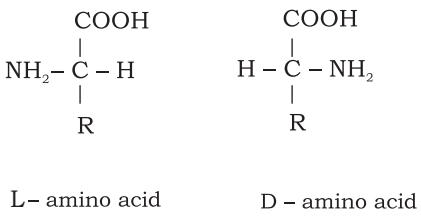


Fig. 3.16: L and D isomers of amino acid.

Table 3.4: 20 standard amino acids along with their three letter and one letter symbol

Amino acid	Abbreviation symbol (Three letter)	Single letter symbol
Polar, uncharged R groups		
Serine	Ser	S

© NCERT not to be republished

Threonine	Thr	T
Cysteine	Cys	C
Asparagine	Asn	N
Glutamine	Gln	Q
Aromatic R groups		
Phenylalanine	Phe	F
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Non-polar aliphatic amino acids		
Glycine	Gly	G
Valine	Val	V
Alanine	Ala	A
Proline	Pro	P
Leucine	Leu	L
Isoleucine	Ile	I
Methionine	Met	M
Positively charged (basic) R groups		
Lysine	Lys	K
Arginine	Arg	R
Histidine	His	H
Negatively charged (acidic) R groups		
Aspartate	Asp	D
Glutamate	Glu	E

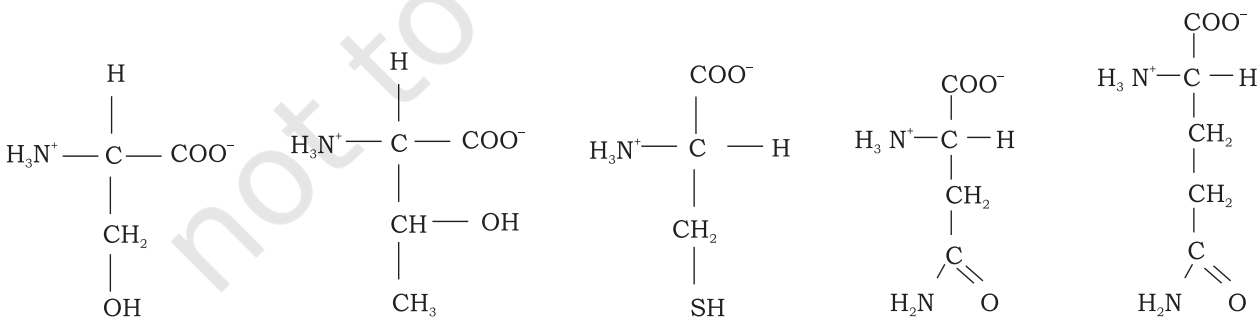


Fig. 3.17: Structure of polar uncharged amino acids

© NCERT not to be republished

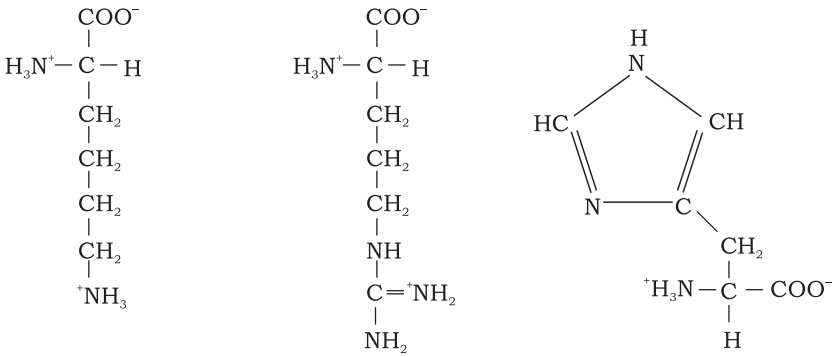
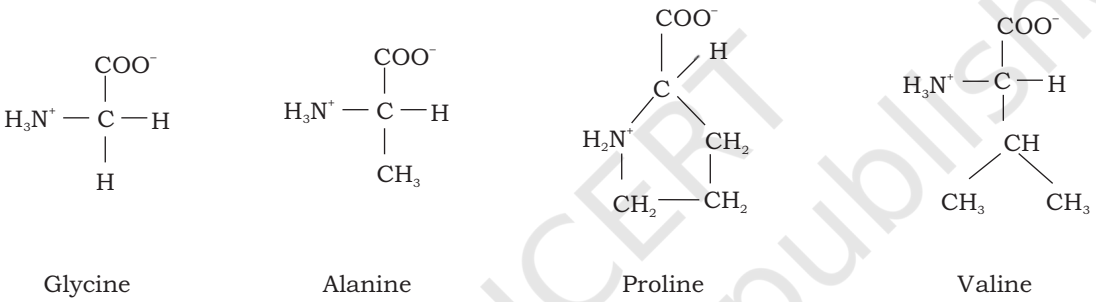


Fig. 3.18: Structure of positively charged amino acids



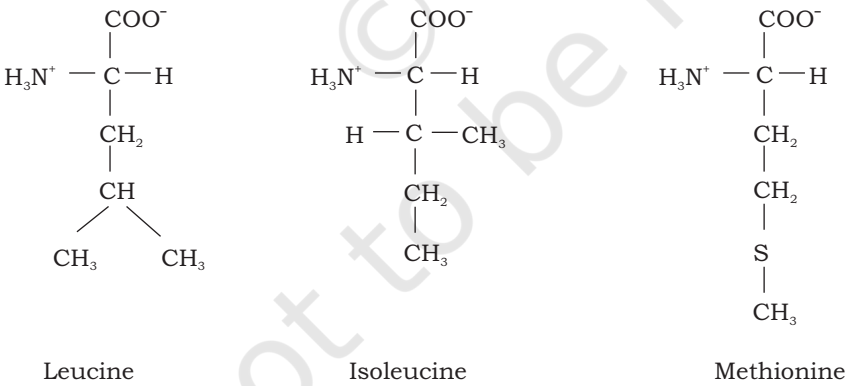


Fig. 3.19: Structure of non-polar aliphatic amino acids

© NCERT not to be republished

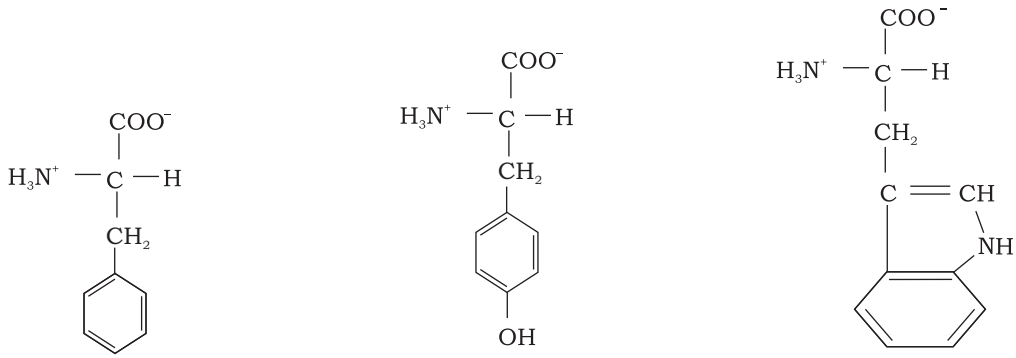


Fig. 3.20: Structure of aromatic amino acids

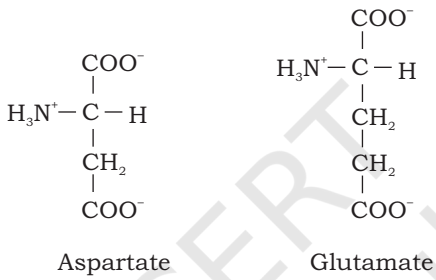


Fig. 3.21: Structure of negatively charged amino acids

Nomenclature

The numbering of carbon in amino acids is done using Greek letters. The additional carbons attached to α-carbon in an R group are designated β (beta), γ (gamma), δ (delta), ε (epsilon) and so on proceeding out from α-carbon. Like other organic molecules, numbering of carbon atoms begins from the carboxylic group and carboxylic carbon being C-1 and α-carbon would be C-2 (Fig. 3.22).

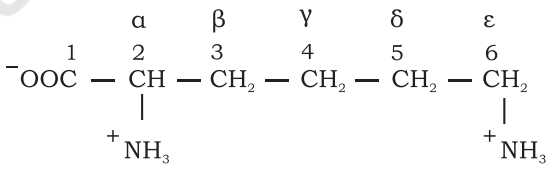


Fig. 3.22: Numbering of carbon atoms of L-lysine

Electrochemical properties of amino acids

At physiological pH (pH=7) the α-COOH group and α-NH_2 group of amino acids are ionized (deprotonated) in solutions

© NCERT not to be republished

to form $-COO^-$ (bearing negative charge) and the $-{}^+NH_3$ (bearing a positive charge), respectively. This dipolar state of amino acids is called **zwitterion.** Zwitterions state occurs as a result of migration of proton from carboxyl group to amino group (Fig. 3.23).

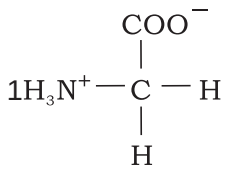


Fig. 3.23: Zwitterion of glycine

Non-standard and non-protein amino acids

Some amino acids are non-standard amino acids, which occur naturally in the cells but do not take part in protein synthesis. They are generated after protein synthesis by modification of the specific standard amino acids. Examples are 4-hydroxyproline (formed by hydroxylation of proline), 5-hydroxylysine, selenocysteine and γ-carboxyglutamic acid. Amino acids that are not part of proteins are widely present in various plants, animals, and microbes known as non-protein amino acids. Some non-protein amino acids are L-ornithine, L-citrulline, β-alanine, creatine, and γ-aminobutyrate.

3.4 PROTEIN STRUCTURE

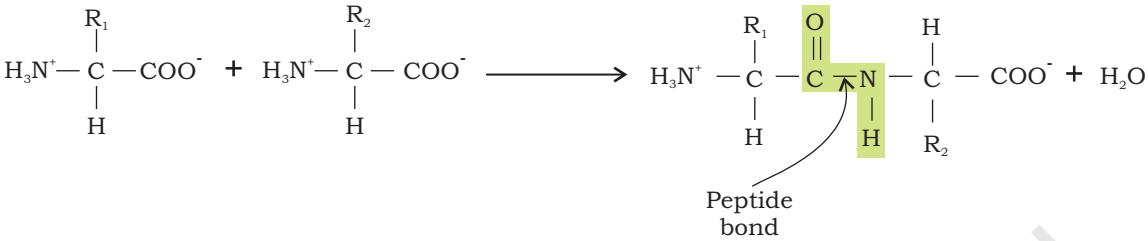
Proteins are one of the macromolecules present in all cells from simplest bacteria to human beings and plants. A single cell may contain thousands of different proteins. Proteins are made up of 20 naturally occurring amino acids, which are linked covalently to form a linear sequence. The most remarkable thing is that cells can produce proteins with entirely different structures and properties by joining the amino acids in many different combinations. From these building blocks, the organisms can make vast variety of products such as catalysts (enzymes), antibodies, transporters, hormones, transcription factors, muscle fibres, membrane proteins etc. These products participate in all vital processes such as, in the transport of oxygen and nutrients to each cell, muscle contraction, transmission of nerve impulse, control of metabolism, growth and differentiation, providing mechanical support, immune response, signal transduction and many more.

Composition of amino acids and their order in proteins decide the structure of a protein. Four levels of protein structures exist namely primary, secondary, tertiary, and quaternary, which are described below.

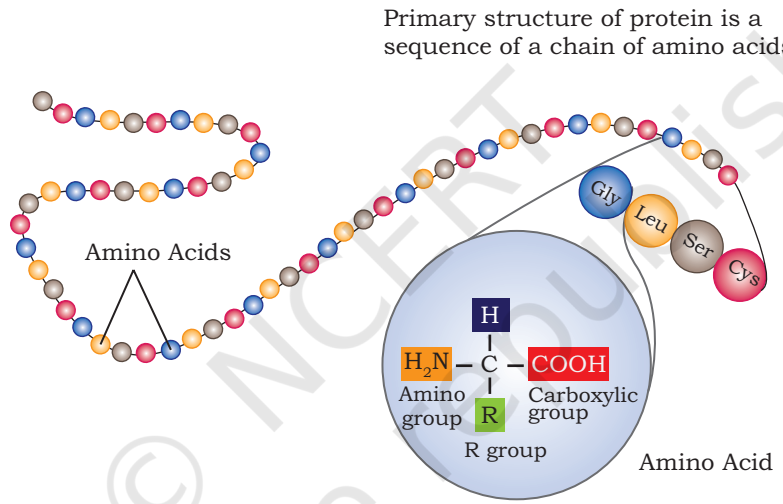
© NCERT not to be republished

3.4.1 Primary structure of proteins

The primary structure of proteins is the linear chain of amino acid sequences that are linked through peptide bonds (Fig. 3.24).



(a) Reaction showing peptide bond formation



(b) Long polypeptide chain containing linear sequence of amino acids linked by peptide bond

Fig. 3.24: Primary structure of protein (a) Reaction showing peptide bond formation (b) Long polypeptide chain containing linear sequence of amino acids linked by peptide bond

The peptide bond (also called amide bond) is formed by coupling of the α-carboxyl group of one amino acid to an α-amino group of another amino acid. Formation of a peptide bond between two amino acids is accompanied by the loss of water molecules. Each amino acid unit in a polypeptide is therefore called a residue.

On the basis of number of amino acids constituting a chain, the peptides may be called as a dipeptide (containing two amino acid units), a tripeptide (containing three amino

© NCERT not to be republished

acid units) and so on. A polypeptide chain has polarity because its ends are different. The two ends are named as amino (or N-terminal) and carboxy (or C-terminal) ends. In the naming of a polypeptide, the convention is that the N-terminal residue, present at the extreme left, is written first and the C-terminal residue is written at the end. Thus, in the pentapeptide Tyr-Ala-Gly-Ser-Leu (YAGSL), tyrosine is the amino-terminal residue and leucine is the carboxy-terminal residue.

In a polypeptide, constituent amino acids are named by adding the suffix –yl (because all these are acyl groups) to all the amino acids except the last one where no suffix is added. For example, the pentapeptide Tyr-Ala-Gly-Ser-Leu (YAGSL) is named as Tyrosyl-L-alanyl-L-glycyl-L-seryl-L-leucine.

On an average, the majority of natural polypeptides contain 50 to 2,000 amino acid residues. The mean molecular weight of an amino acid residue is about 110 Da (Dalton is the unit of mass).

© NCERT not to be republished

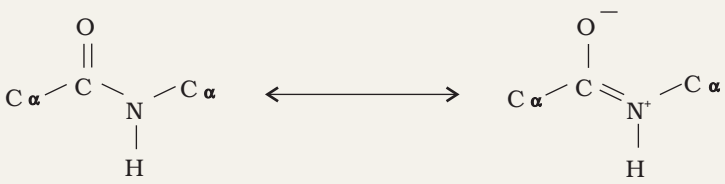
Box 2

Conformation of peptide bond

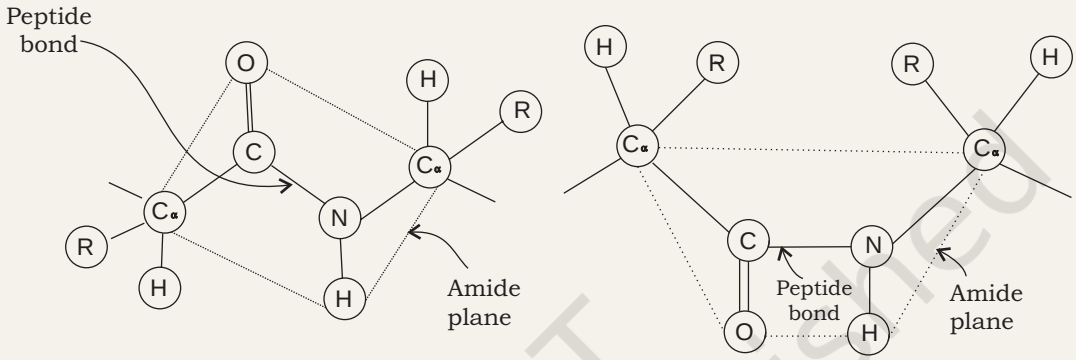
Six atoms of a pair of amino acids linked by a peptide bond, lie in the same plane i.e., has a planar structure. These atoms are the α-carbon atom and CO group from the first amino acid, and NH group and α-carbon from the second amino acid. Planar structure is the result of resonance interactions that give the peptide bond an ~40% double bond character. Due to this double bond character, the peptide bond cannot rotate freely.

The flexibility of polypeptide backbone

The backbone of protein is a linked sequence of rigid planar peptide bonds. In contrast to peptide bond, which has partial double bond character, the bonds between the – NH_2 group and α-C, and between α-C and –COOH group are pure single bonds. The rotation of two bonds, adjacent to peptide bond results in various orientations. Because of this freedom of rotation about two bonds of each amino acids residues linked, there are many possible ways proteins can fold. The rotation angles of these bonds can be specified by dihedral angles (torsion angles). The rotation angle about the Cα–N bond is called phi (φ). The angle of rotation between Cα–C bond is called psi (ψ). A clockwise rotation about either bond as viewed from the front of the back group corresponds to a positive value. The φ and ψ angles determine the path of the polypeptide chain. By convention, both φ and ψ are defined as



Resonance structure of peptide bond

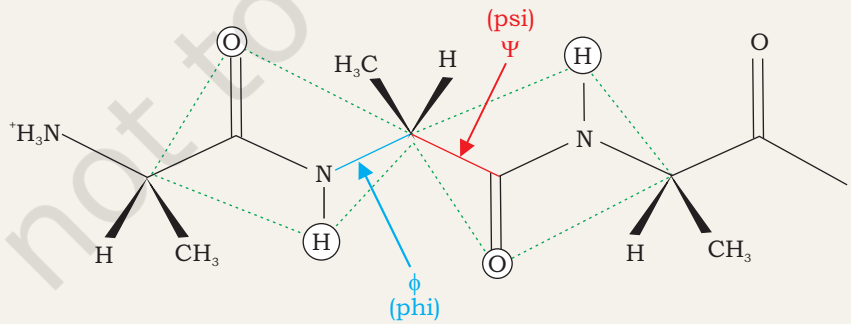


Trans and cis conformation of a peptide bond

0° in the conformation, in which the two peptide bonds are connected to a single carbon in the same plane. The dihedral angles, φ and ψ can have any value between -180° and +180°, but most of the values of φ and ψ are prohibited. The conformation in which φ and ψ are both 0° is also prohibited because of steric interference between atoms of polypeptide backbone and the amino acid side chains.

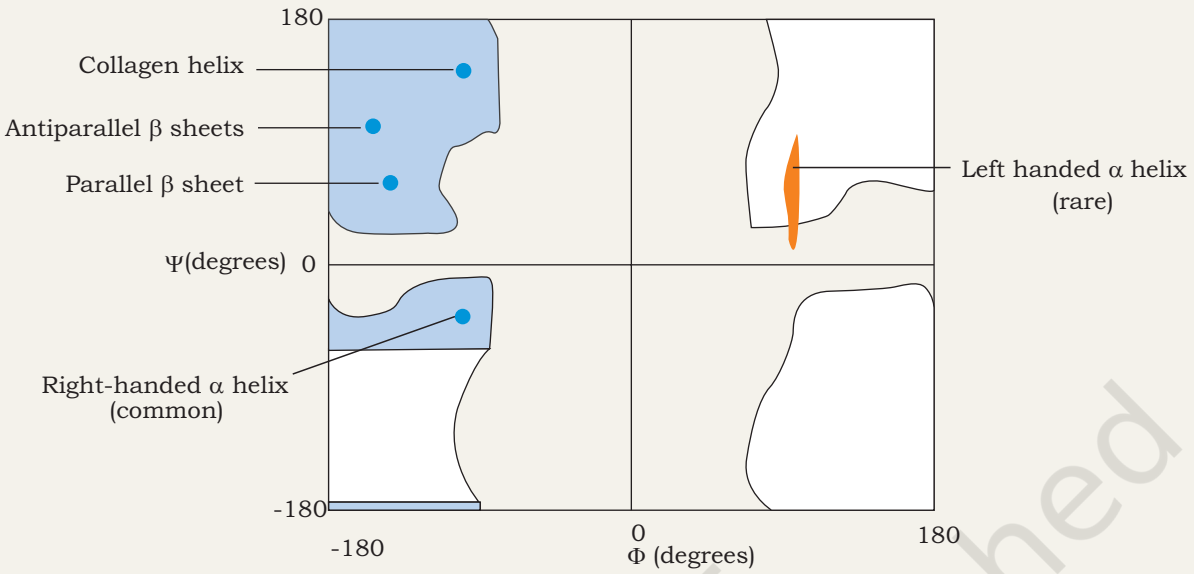
Ramachandran Plot

The conformation of a fully stretched polypeptide chain having φ = ψ = 180° can be determined if the value of φ and ψ for each amino acid residue is present in the polypeptide chain. In 1963, G.N. Ramachandran identified that an amino acid residue in polypeptide chain cannot just have any pair of values of φ and ψ. The values of φ



Rotation about bonds in the polypeptide. φ is rotation angle between Ca–N bond and ψ is rotation angle between Ca–C bond

© NCERT not to be republished



A Ramachandran Plot

and ψ that are permissible can be predicted by assuming atoms as hard/solid spheres. The allowed values of φ and ψ can be represented on a two-dimensional plot called Ramachandran plot. For poly-L-serine, this plot shows three separate allowed regions (shaded in the Figure). One region contains φ-ψ values that produce the parallel β sheet, the antiparallel β sheet, and the collagen helix. The second region includes φ-ψ values that produce the right-handed α helix, and the third region has left-handed α helix. Though left-handed α helices come in sterically allowed region, but this conformation is not found in proteins due to less stability.

3.4.2 Secondary structure of proteins

Secondary structure of protein deals with folding of the polypeptide chain. In 1951, Linus Pauling and Robert Corey proposed two types of periodic structures called α helix and β pleated sheet.

α Helix

A polypeptide chain with planar peptide bonds can form a helical structure by twisting about the C_α–N and the C_α–C bonds called α-helix. Thus, an α-helix is a rod-like structure (Fig. 3.25). The hydrogen bonds between NH and CO group of the main chain stabilise the helix. CO group of each amino acid forms hydrogen bonds with the NH group of the amino acid that is situated four residues ahead in

© NCERT not to be republished

the sequence. All the main chain NH and CO groups are hydrogen bonded, except the amino acids near the ends of the helix. Also, α helices are connected by loops.

The α helix can be right-handed (clockwise) or left-handed (counter clockwise). Right-handed helices are sterically more stable due to less clash between side chains and the backbone. All known polypeptides contain right-handed α helix. The occurrence of α helical content in proteins ranges widely. For example ferritin, a blood protein that helps storage of iron, has 75% of its amino acid residues and form α helix.

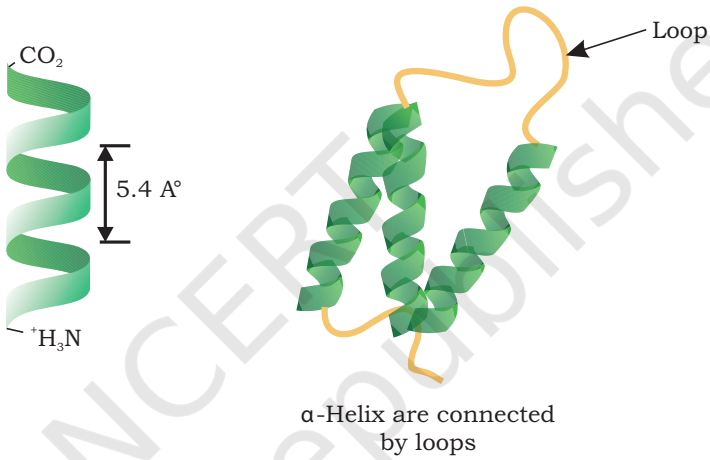


Fig. 3.25: α-Helix

These α-helical structures, present in myosin and tropomyosin of muscles and keratin of hair, provides mechanical strength to the stiff bundles of fibers.

β Pleated sheet

The second type of periodic structure that Pauling and Corey suggested was β pleated sheet. In contrast to α helix, β pleated sheet involves hydrogen bonds between groups from residues distant from each other in the linear sequence. In β sheets, two or more strands widely separated in the backbone of the polypeptide chain are arranged side by side, with hydrogen bonds between the strands. Based on the orientation of the strands, β sheets are of two types: parallel and anti-parallel β sheets

© NCERT not to be republished

(Fig. 3.26). If the strands run in the same direction they are called parallel β sheets; if strands run in opposite direction they are called anti-parallel β sheets. Parallel β sheets are less twisted as compared to anti-parallel.

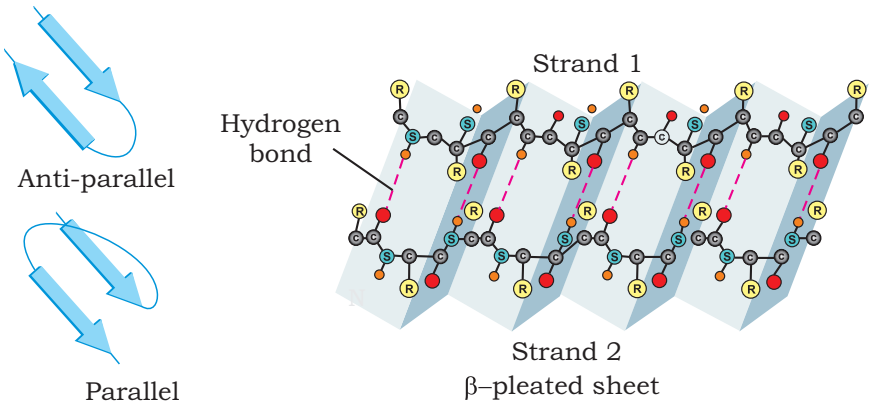


Fig. 3.26: Structure of β pleated sheets

3.4.3 Tertiary structure of protein

The overall three-dimensional arrangement of all residues in a protein is called tertiary structure of the protein. In contrast to secondary structure, the tertiary structure includes larger range aspects of amino acid sequences. The amino acids that are far apart in a polypeptide chain within different types of secondary structures may interact with each other to form the completely folded structure of a protein. In tertiary structure, there is an involvement of some additional bonds like disulfide, hydrogen, hydrophobic and ionic (Fig. 3.27). The disulfide bonds are formed by oxidation of a pair of cysteine residues. These bonds make the protein globular in shape. Some enzymes, transport protein, peptide hormones, and immunoglobulins are globular in shape. In the tertiary structure, the polar R groups of amino acid residues are located on the outer side because of their hydrophilic nature, and the non-polar R groups of amino acid residues are located in the interior to form hydrophobic interaction. The three-dimensional structure of many proteins are determined by X-ray crystallography and nuclear magnetic resonance (NMR) techniques, etc.

© NCERT not to be republished

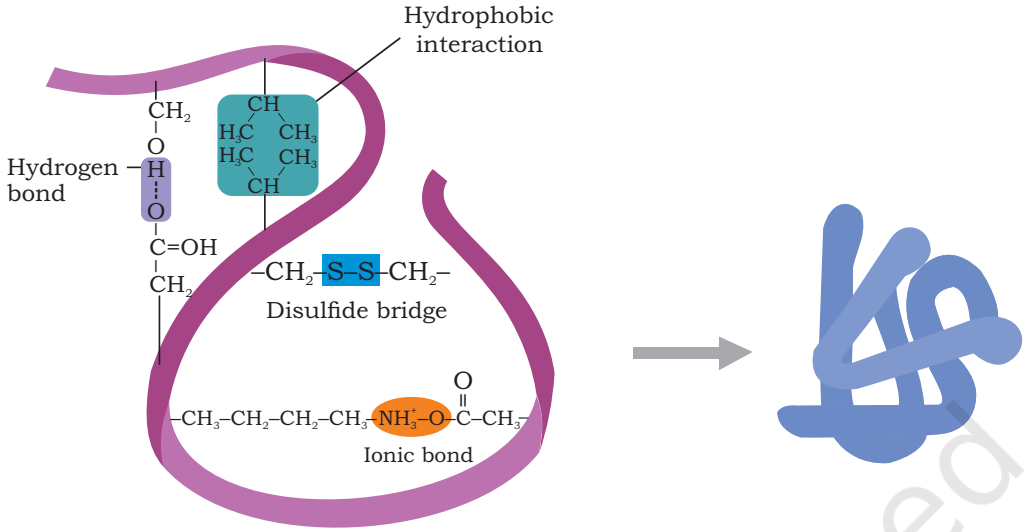


Fig. 3.27: Tertiary structure of protein

3.4.4 Quaternary structure of protein

Some proteins may consist of more than one polypeptide subunit, which may be identical or different in their primary structure. These subunits associate specifically to each other to form a comparatively larger and complex molecule, i.e., quaternary structure. Quaternary structure is the spatial arrangement of the protein subunits. These subunits are stabilised by hydrogen bonds, electrostatic interactions, ionic bonds and disulfide bridges. Depending upon subunit numbers, the quarternary structures may be dimer, trimer, etc. Identical protein subunits form homodimer, homotrimer, etc. whereas non-identical subunits form heterodimers, heterotrimer, etc. The diagrammatic representation of the

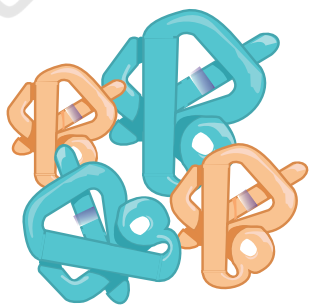


Fig. 3.28: The quaternary structure a protein

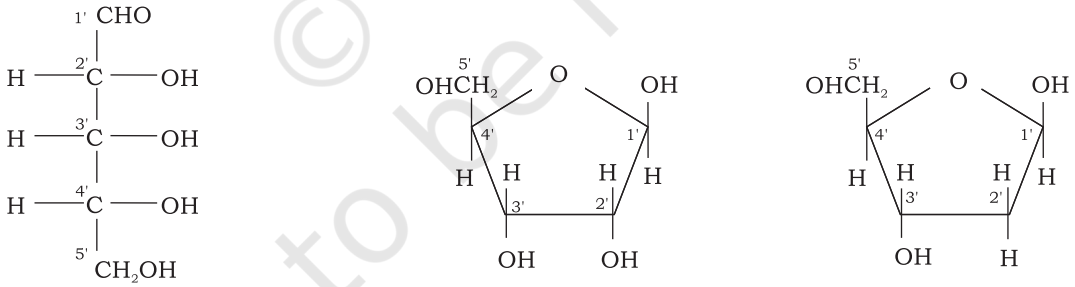
© NCERT not to be republished

quaternary structure of a tetramatic protein consisting of two identical subunits each is given in (Fig. 3.28).

3.5 NUCLEIC ACIDS

The cell organelles, namely nucleus, mitochondria and chloroplast contain nucleic acid within them. Within nucleus, nucleic acids are associated with histone proteins to form chromatin. Nucleic acids are polymers of nucleotides linked through phosphodiester linkages. Two types of nucleic acids are present in cells, namely deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA acts as a genetic material and inherits the information from one generation to next. However, RNA serves as genetic material in some viruses.

Nucleotides of DNA and RNA are made up of nitrogenous base, sugar, and phosphate. The sugar that is present in nucleic acids is pentose sugar which is of two types; the one present in DNA is **2'-deoxy-D-ribose,** and the other present in RNA is **D-ribose**. Both the pentoses are present as closed five membered rings (Fig. 3.29). For numbering pentoses of nucleotides the carbon numbers are given a prime (′) designation to differentiate them from the numbered atoms of the nitrogenous bases.



D-ribose sugar (a pentosugar) D-ribose 2'-deoxy-D-ribose

Fig. 3.29: Structure of pentose sugars present in nucleic acids

Nitrogenous bases are of two types; **purines** and **pyrimidines**. The two purine bases of DNA and RNA are **adenine (A)** and **guanine (G)**. Among pyrimidines, **cytosine (C)** is present in both DNA and RNA, **thymine (T)** is present in DNA only and

© NCERT not to be republished

uracil (U) is present in RNA only. The structure of five major bases is shown in (Fig. 3.30). The purine and pyrimidine bases contain aromatic ring structures which absorb light at a wavelength near 260 nm.

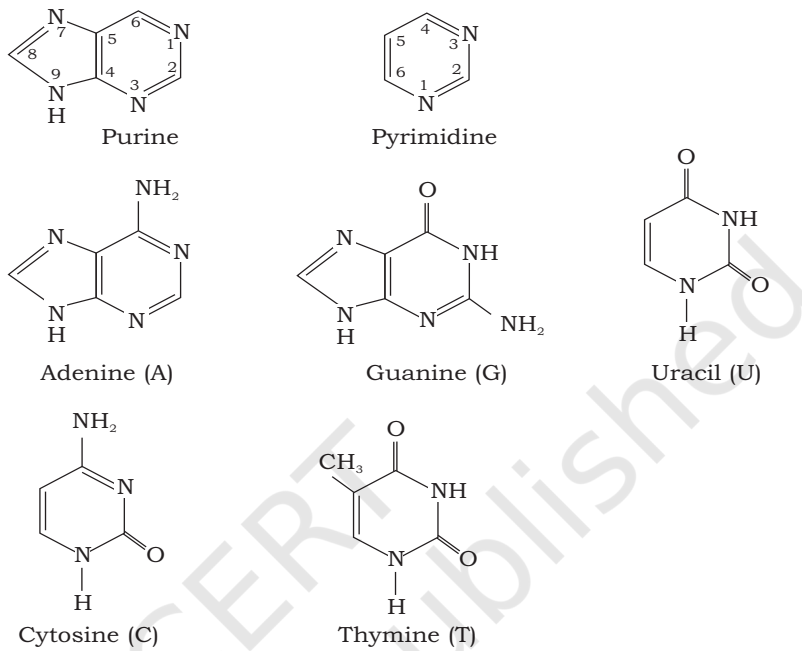


Fig. 3.30: Structure of nitrogenous bases present in nucleic acids

A base (either purine or pyrimidine), a pentose sugar unit and a phosphate group are linked together to form a **nucleotide**. On the other hand, a base linked only to a pentose sugar unit without a phosphate group is called a **nucleoside**. The four nucleosides of DNA are named as deoxyadenosine, deoxyguanosine, deoxycytidine, and deoxythymidine. The nucleotides in DNA are of four types, namely deoxyadenylate (deoxyadenosine-5'-monophosphate; dAMP), deoxyguanylate (deoxyguanosine-5'-monophosphate; dGMP), deoxycytidylate (deoxycytidine-5'-monophosphate; dCMP) and deoxythymidylate (deoxythymidine-5'-monophosphate; dTMP) (Fig. 3.31). Nucleotides in DNA are indicated by prefix 'd' as they contain deoxyribose, rather than ribose sugar.

The four nucleosides of RNA are called adenosine, guanosine, cytidine, and uridine, which, when bound to a phosphate group, make ribonucleotides or ribonucleoside-5'-monophosphate. There are four types of ribonucleotides namely; adenylate (adenosine-5'-monophosphate; AMP), guanylate (guanosine-5'-monophosphate; GMP), cytidylate

© NCERT not to be republished

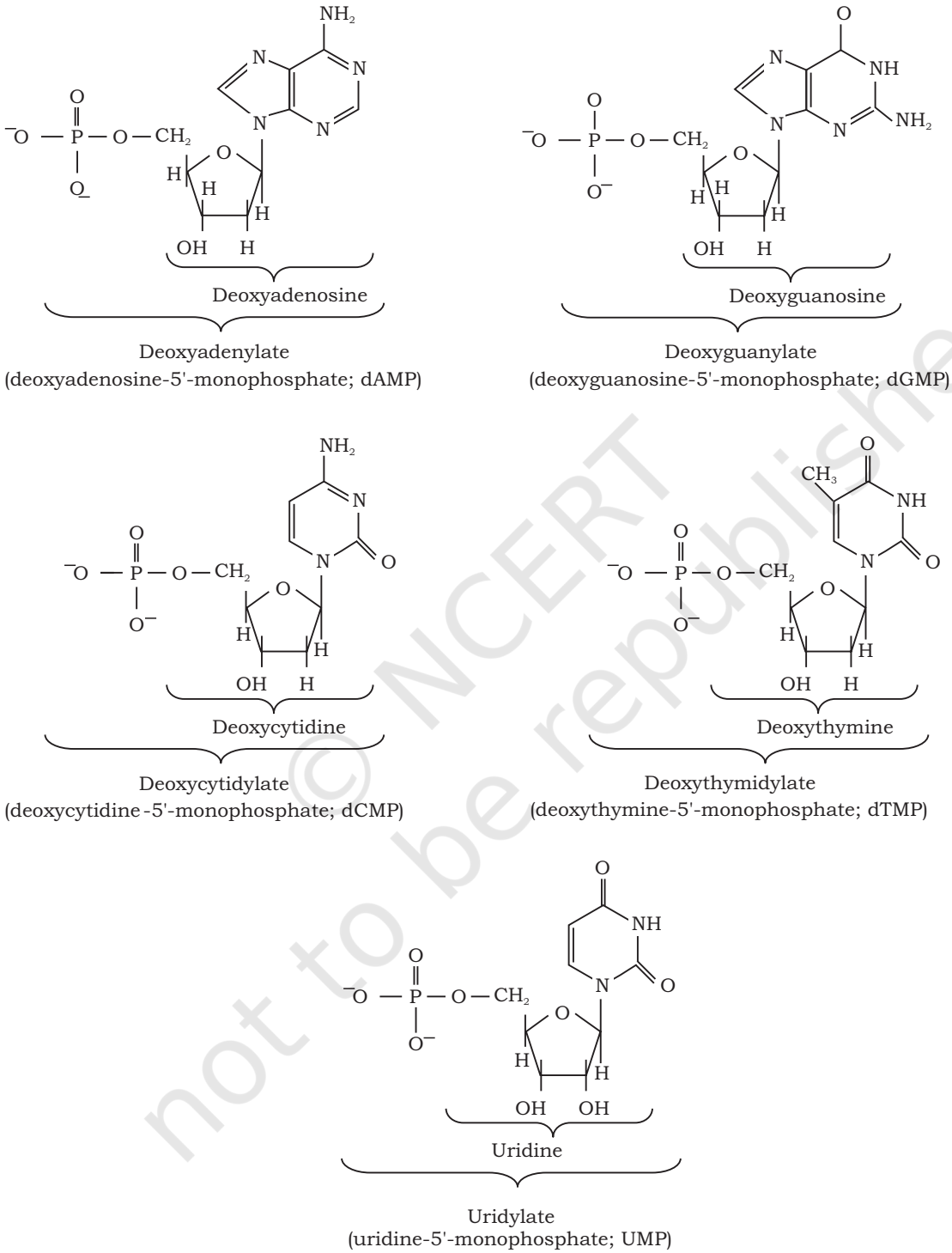


Fig. 3.31: The nucleosides and nucleotides present in DNA and RNA

© NCERT not to be republished

(cytidine-5'-monophosphate; CMP), and uridylate (uridine-5'-monophosphate; UMP) (Fig. 3.31).

3.5.1 Polynucleotide chain

Nucleotides of both DNA and RNA are covalently linked, in which the 3' hydroxyl (-OH) group of the sugar of one nucleotide unit is esterified to -OH group of phosphate attached to 5' carbon atom of the sugar of the next nucleotide, forming a phosphodiester linkage (Fig. 3.32). Like a polypeptide chain, a polynucleotide chain has a specific polarity, i.e., distinct 5' and 3' ends. The 5' end has phosphate group at C-5' of sugar, while 3' end has free -OH group at C-3' of ribose. The base sequence in a polynucleotide chain is written in the 5'→3' direction.

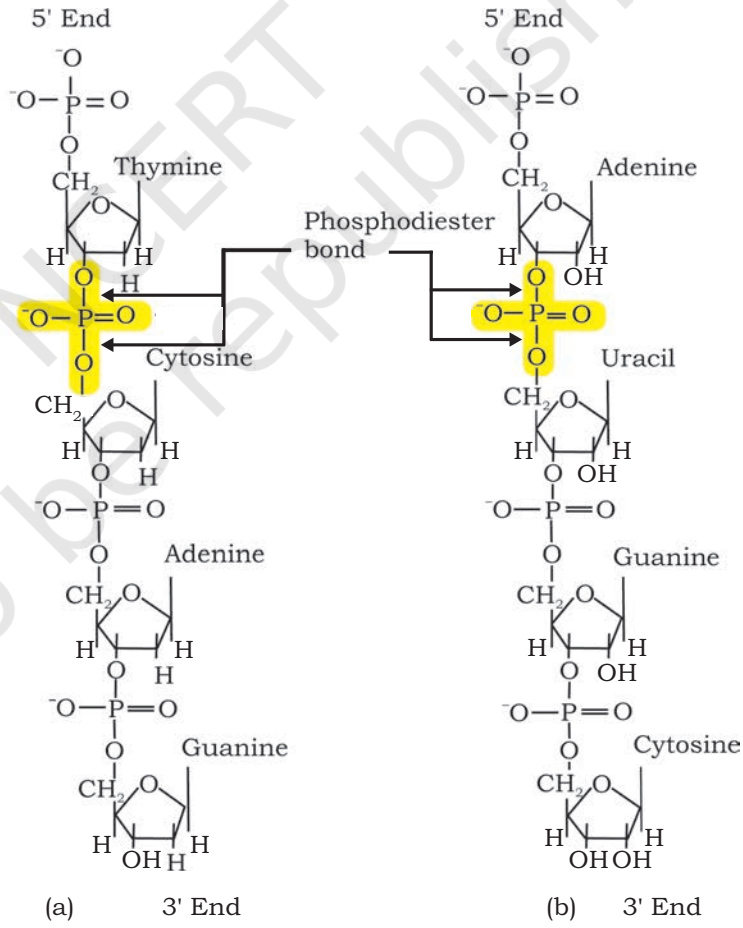


Fig. 3.32: Primary structure of (a) DNA and (b) RNA

© NCERT not to be republished

James Watson and Francis Crick in 1953 proposed the double helical (three dimensional) structure of DNA [Fig. 3.33 (b)]. It consists of two polynucleotide chains of DNA wound around the same axis to form right handed double helix. The two strands are oriented antiparallel, i.e., their 3' and 5' phosphodiester bonds run in opposite directions. The sugar and phosphate form the backbones of the double helix and remain exposed to the polar environment. The bases of both the strands are stacked inside the core of the double helix and make it hydrophobic. Within the helix, each nucleotide base of one strand makes hydrogen bonds in the same plane with the base of the other strand. A of one strand forms two hydrogen bonds with T of the other strand (A=T) and *vice versa* and G on one strand forms three hydrogen bonds with C on the other strand (G≡C) and *vice versa* [Fig. 3.33 (a)].

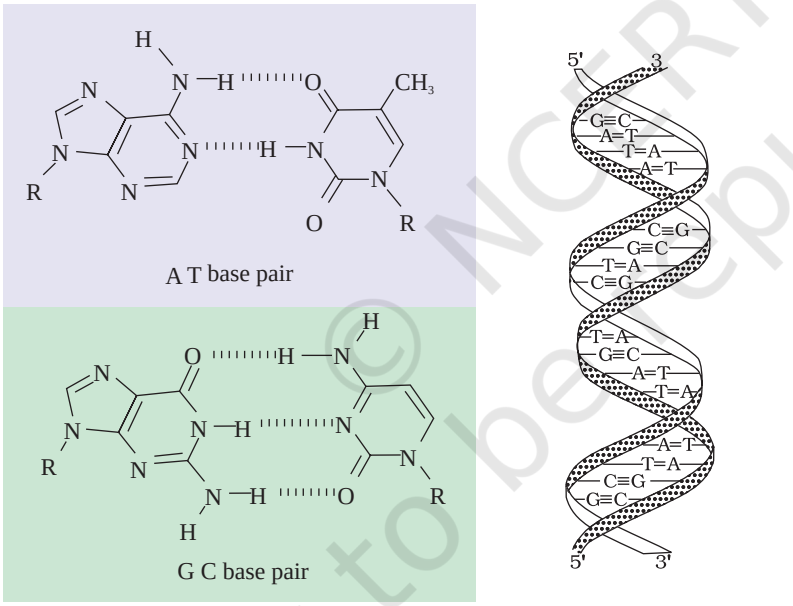


Fig 3.33: (a) Base pairing in DNA and (b) Watson-Crick double helical model of DNA

The Watson-Crick structure of DNA is also known as B-DNA. The B-DNA is the most stable form of DNA. The other two structural variants are A-DNA and Z-DNA. The A-DNA is right handed double helix. It is wider and contains 11 base pairs per helix. However, Z-DNA is left handed double helix with 12 base pairs per helix.

© NCERT not to be republished

The bases of two strands of DNA double helix are joined by hydrogen bonds [Fig. 3.33 (a)]. Heating disrupts hydrogen bonds between the base pairs, and thereby separation of the two strands of DNA takes place called **denaturation** or **melting**. The temperature at which half of the DNA is denatured is called **melting temperature (T_m).** Apart from heating, melting is also caused by acid or alkali. The two strands of nucleic acids that are separated by melting, can spontaneously reassociate to form a double helix if temperature is decreased below T_m. This process of reassociation or renaturation is called **annealing**.

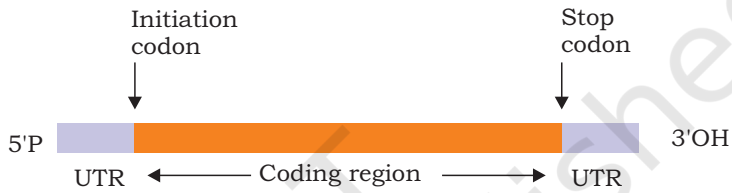


Fig 3.34: Schematic representation of a typical prokaryotic mRNA

3.5.2 Types of RNA

Messenger RNA (mRNA)

It is a single strand linear polyribonucleotide chain that carries genetic information from DNA to ribosome. At the 5' end is UTR (untranslated region) containing no genetic information for polypeptide synthesis. It is followed by initiation codon, coding region, and stop codon. At the 3' end is present another UTR (Fig. 3.34). In eukaryotic mRNA, 5' end contains guanylate which is methylated at its N-7. This process is called capping. 3' end of mRNA also undergoes polyadenylation (addition of several adenylate residues).

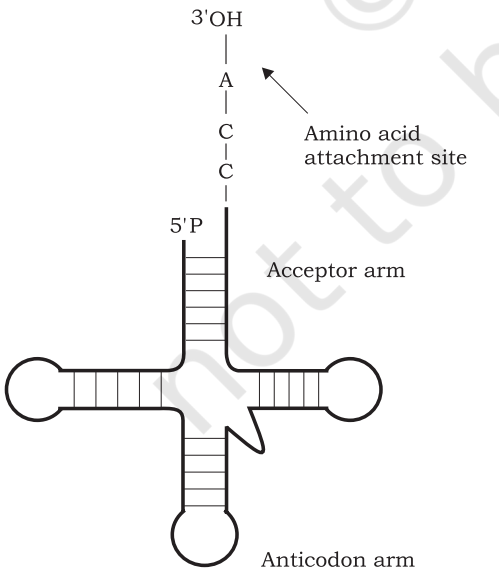


Fig. 3.35: Structure of a tRNA molecule

Ribosomal RNA (rRNA)

It forms the structural components of ribosomes. 70S prokaryotic ribosomes have 16S rRNA in smaller subunit (30S), 23S and 5S rRNAs in larger subunit (50S). The eukaryotic ribosome (80S) has 18S rRNA in smaller subunit (40S) and 28S, 5.8S and 5S rRNAs in larger subunit (60S).

© NCERT not to be republished

Transfer RNA (tRNA)

These are small RNA molecules that transfer amino acids to ribosome during protein synthesis. It is a single polyribonucleotide chain folded to form four arms having a clover-leaf secondary structure (Fig. 3.35). The acceptor arm has a CCA sequence at 3'OH end, which is amino acid binding site. The anticodon arm has anticodon, a set of three bases which recognises a specific codon of mRNA during translation.

SUMMARY

- Carbohydrates are made up of carbon, hydrogen, and oxygen atoms, and are widely distributed in animal and plant tissues.
- Carbohydrates are categorised into 4 major classes, namely monosaccharides (e.g., glucose, ribose), disaccharides (e.g., sucrose, lactose), oligosaccharides (e.g., raffinose), and polysaccharides (e.g., starch, glycogen).
- Monosaccharides can be classified as trioses, tetroses, pentoses, hexoses, etc., depending on the number of carbon atoms in the molecule.
- Pentose sugar such as ribose is an important component of nucleic acids and coenzymes.
- Two monosaccharides linked by glycosidic bond make a disaccharide (e.g., sucrose).
- Polysaccharides perform various roles as the storage form of energy and also as structural component.
- Lipids are organic compounds found in living organisms. These are made up of hydrophobic fatty acid chains linked to glycerol via ester bond.
- Fatty acids are long chain hydrocarbon containing carboxylic acid group. Fatty acids may be saturated (having no double bonds) or unsaturated (having one or more double bonds). Nomenclature of fatty acids is based on total number of carbons, total number of double bonds, and position of double bonds.
- Lipids are broadly classified into two classes– simple and compound lipids.
- Simple lipids include triacylglycerol and waxes, which are esters of glycerol with fatty acids and esters of high molecular weight alcohol with fatty acids, respectively.

© NCERT not to be republished

- Compound lipids include membrane lipids which are of two types, glycerophospholipids and sphingolipids. Compound lipids are amphipathic molecules made up of hydrophilic phosphate tails.
- Glycerophospholipid contains glycerol backbone and sphingolipids contain sphingoid base backbone.
- Steroids are another class of compound lipids made up of four-fused ring structure. Cholesterol is the most common steroid found in animals. It is the precursor of all steroid hormones and vitamin D.
- The amino acids are organic compounds containing amine ($-NH_2$) and carboxyl (-COOH) functional groups, along with a side chain (R group) specific to each amino acid.
- There are 20 standard amino acids and some non-standard amino acids (e.g. 4-hydroxy proline, 5-hydroxy lysine, etc.) and some non-protein amino acids (e.g., L-ornithine, L-citrulline, etc.)
- In a polypeptide chain, the amino acids are linked covalently via peptide bond in a linear fashion to make a protein.
- There are four levels of protein structures, namely primary, secondary, tertiary, and quaternary.
- Primary structure of proteins is the linear chain of amino acid sequences linked through peptide bonds.
- Secondary structure of protein is the three dimensional form of polypeptide.
- The two major types of secondary structures are α-helix and β-sheets.
- Tertiary structure is the three dimensional arrangement of protein.
- Quaternary structure is the complex arrangement of folded protein subunits, stabilised through hydrogen bonds, electrostatic interactions, etc.
- Nucleic acids are the polymer of nucleotide which contain nitrogenous bases (adenine, guanine, cytosine, thymine and uracil), sugar and phosphate.
- Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are the two types of nucleic acids.
- DNA is genetic material of almost all the organisms except some viruses where RNA is the genetic material.
- DNA contains deoxyribose sugar and thymine while RNA contains ribose sugar and uracil in place of thymine.

© NCERT not to be republished

- In DNA, adenine forms 2 hydrogen bonds with thymine (A=T) and cytosine forms 3 hydrogen bonds with guanine (C≡G).
- J.Watson and F.Crick (1953) described the three dimensional double helical structure of DNA.
- The mRNA, tRNA, rRNA are the major types of RNA.

EXERCISES

1. Describe the classification of carbohydrates.
2. Differentiate between D- and L-forms of glucose.
3. Draw the structure of a disaccharide made up of two monosaccharides glucose and fructose.
4. Draw the partial structure of starch and glycogen.
5. Write the major functions of carbohydrates.
6. Describe isomerisation in monosaccharides.
7. Differentiate between sphingolipids and glycerolipids.
8. Why are membrane lipids called amphipathic?
9. Differentiate between saturated and unsaturated fatty acids.
10. Describe the various categories of amino acids.
11. What is zwitterion? Draw the structure of zwitterion.
12. What are non-standard and non-protein amino acids?
13. How the peptide bonds are formed?
14. Draw the structure of Lys-Glu-Lys.
15. Describe the various secondary structures of protein.
16. Differentiate between tertiary and quaternary structure of proteins.
17. Differentiate between nucleosides and nucleotides.
18. Explain the primary structure of DNA.
19. Draw the structure of A-T-C-G oligonucleotide.
20. Explain Watson and Crick model of DNA.
21. Describe the various forms of DNA.
22. Describe the clover leaf model of tRNA.

© NCERT not to be republished

23. In carbohydrates which of these functional groups are present
    (a) Alcohol and carboxyl groups
    (b) Aldehyde and ketone groups
    (c) Hydroxyl and hydrogen groups
    (d) Ether and ester groups

24. Which of the following is a non-reducing disaccharide?
    (a) Maltose
    (b) Lactose
    (c) Sucrose
    (d) Cellobiose

25. The repeating units of proteins are
    (a) Glucose units
    (b) Amino acids
    (c) Fatty acids
    (d) Nucleotides

26. Which of the following is the most common secondary structure of proteins
    (a) α-helix
    (b) β-pleated sheet
    (c) Both (a) and (b)
    (d) None of the above

27. A nucleotide contains
    (a) Nitrogenous base, sugar and phosphate
    (b) Sugar and phosphate
    (c) Nitrogenous base and sugar
    (d) None of the above

28. The two strands in DNA double helix are joined by
    (a) Covalent bond
    (b) Hydrogen bond
    (c) Glycosidic bond
    (d) Phosphodiester bond

29. Which is an example of storage lipid?
    (a) Fatty acids
    (b) Triacylglycerol
    (c) Sphingolipids
    (d) Eicosanoids

30. In glycerolipids fatty acids are joined to glycerol through which bond?
    (a) Phosphodiester bond
    (b) Glycosidic bond
    (c) Peptide bond
    (d) Ester bond

© NCERT not to be republished

11150CH04

Chapter 4

Enzymes and Bioenergetics

4.1. *Enzymes: Classification and Mode of Action*

4.2. *Brief Introduction to Bioenergetics*

4.1 Enzymes: Classification and Mode of Action

Enzymes are biocatalysts which catalyse the biochemical reactions both *in vivo* as well as *in vitro*. These are highly specific to substrates and have great catalytic power, i.e., they enhance the rate of reaction tremendously without being changed. All enzymes are proteins with exception of some small group of catalytic RNA molecules called **ribozymes**. Like proteins, the molecular weight of enzymes ranges from about 2000 to more than one million Dalton. Enzymatic activity of proteinaceous enzymes may be affected depending on the conformational structure as well as its denaturation. There are many enzymes which require cofactors for their catalytic activity. The cofactor may be a complex organic molecule called coenzyme (Table 4.1) or it may be a metal ion such as Fe^{2+}, Mn^{2+}, Zn^{2+}, Mg^{2+} (Table 4.2). An enzyme plus its cofactor is called **holoenzyme**. In such cases, the protein component in cofactor requiring enzyme is called **apoenzyme**.

© NCERT not to be republished

Table 4.1: Some coenzymes and their precursor vitamins and their role

Coenzyme	Precursor vitamin	Role in the catalytic reaction
Biocytin	Biotin (vitamin B7)	Transfer of CO_2
Coenzyme B12 (5'-adenosylcobalamin)	Vitamin B12	Transfer of an alkyl group
Flavin adenine dinucleotide (FAD)	Riboflavin (vitamin B2)	Transfer of electrons
Coenzyme A	Pantothenic acid (vitamin B5)	Transfer of acyl and alkyl group
Nicotinamide adenine dinucleotide (NAD)	Niacin (vitamin B3)	Transfer of hydride ($:H^-$)
Pyridoxal phosphate	Pyridoxine (vitamin B6)	Transfer of amino group
Thiamine pyrophosphate	Thiamine (vitamin B1)	Transfer of aldehydes
Tetrahydrofolate	Folic acid (vitamin B9)	Transfer of one carbon group

Coenzymes take part in catalysis transiently and are carriers of specific functional groups. Most of the coenzymes are derived from vitamins (organic nutrients required in small amounts in diet).

Table 4.2: Metal ions that serve as cofactors for enzymes

Metal Ions	Enzyme name
Fe^{2+} or Fe^{3+}	Catalase, peroxidase, cytochrome oxidase
Cu^{2+}	Cytochrome oxidase
Mg^{2+}	DNA polymerase
Mn^{2+}	Arginase
K^+	Pyruvate kinase
Mo^{2+}	Nitrogenase, nitrate reductase
Zn^{2+}	Carbonic anhydrase, alcohol dehydrogenase
Ni^{2+}	Urease

When a coenzyme or metal ion is tightly bound through covalent bond with the enzyme protein, it is called a **prosthetic group**.

4.1.1 Classification of enzymes

In order to have a systematic study and to avoid ambiguities considering the fact that new enzymes may also be

© NCERT not to be republished

discovered, International Union of Biochemistry (I.U.B.) in 1964 has adopted classification of enzymes depending on the type of reactions they catalyze. According to this commission, all enzymes are classified into 6 major classes, but recently, another class of enzymes namely translocase has been added (Table 4.3).

Table 4.3: Classification of enzymes adopted by I.U.B.

Class No.	Class name	Type of reaction catalyze
1.	Oxidoreductases	Oxidation-reduction reactions (transfer of electrons)
2.	Transferases	Transfer of groups
3.	Hydrolases	Hydrolytic reactions (transfer of functional groups to water)
4.	Lyases	Addition or removal of groups to form double bonds
5.	Isomerases	Transfer of groups within molecules to yield isomeric forms
6.	Ligases	Condensation of two molecules coupled through ATP hydrolysis
7.	Translocase	Transfer of ion/molecules across the membrane

Isozymes

Many enzymes are present in multiple forms (more than one molecular form) in the same species, tissue or even in the same cell. These enzymes are called **isoenzymes** or **isozymes**. Isoenzymes catalyse the same reaction but have different amino acid composition, hence, possess different physicochemical properties. For example, a glycolytic enzyme, hexokinase exists in four isozyme forms in various tissues. Similarly, lactate dehydrogenase (LDH), involved in anaerobic glucose metabolism has five isozyme forms in human.

Enzyme active site

The catalytic reaction performed by enzymes occurs at a particular site on the enzyme. This site is called **active site**, and represents only small part of the total size of the enzyme. The active site is a clearly defined pocket or cleft in the enzyme molecule where the whole or a portion of substrate can fit. Active site has a three-dimensional structure since it consists of portions of a polypeptide chain. Various non-covalent bonds involved in enzyme substrate binding are electrostatic interactions, hydrogen bonds, Van der Waals forces and hydrophobic interactions.

© NCERT not to be republished

Fischer's Lock and Key Model

In 1894, the introduction of **Lock and Key Model** for the substrate and enzyme interaction was proposed by Emil Fischer. According to this model, complementary structural features are present between enzyme and substrate, and the active site is pre-shaped to fit the substrate. The substrate can fit into its complementary site on the enzyme as a key fits into a lock. This results in the formation of an enzyme-substrate complex (Fig. 4.1).

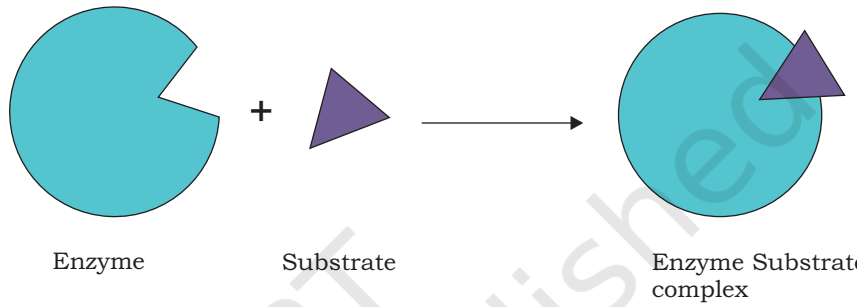


Fig. 4.1: Interaction between an enzyme and its substrate according to lock and key model

Koshland's Induced Fit Model

Daniel Koshland in 1958 proposed **Induced Fit Hypothesis**. He suggested that the structure of a substrate may be complementary to that of the active site in the enzyme-substrate complex but not in the free enzyme. The interaction between the substrate and the enzyme induces conformational changes in the enzyme which aligns the amino acid residues or other groups for substrate binding, catalysis, or both (Fig 4.2). The relationship between a substrate and an active site resembles hand and glove. During interaction, the structure of one component, i.e., substrate or hand remains rigid and the shape of the second

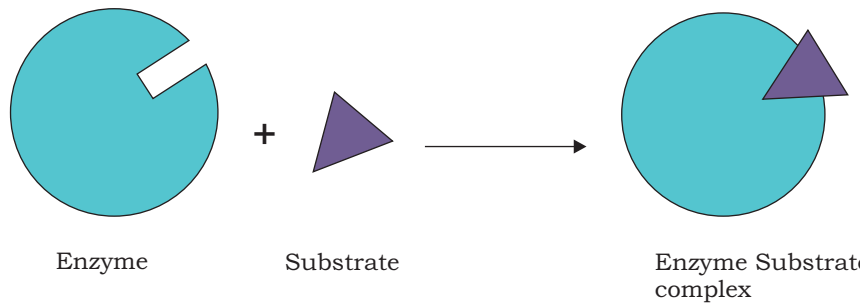


Fig. 4.2: Interaction between an enzyme and its substrate according to induced fit model

© NCERT not to be republished

component, i.e., active site or glove flexible to become complementary to that of the first.

Enzyme specificity

The enzymes are highly specific in action. In fact, the properties that make enzymes such a strong catalyst are their specificity of substrate binding and their ideal arrangement of catalytic groups. Various types of enzyme specificity are: group specificity, absolute specificity, stereospecificity, and geometrical specificity. When enzymes act on several different closely related substrates then it is called **group specificity**. When enzymes act only on one particular substrate, it is called **absolute specificity**. **Stereochemical or optical specificity** occurs when substrate exists in two stereochemical forms (chemically identical but different arrangement of atoms in three-dimensional space) then only one of the isomers will undergo reaction by particular enzyme. For example, D-amino acid oxidase catalyses oxidation of the D-amino acids to keto acids. In **geometrical specificity**, enzymes are specific towards *cis* and *trans* forms. For example, fumarase catalyses the interconversion of fumarate and malate.

4.1.2 Factors affecting enzyme activity

Rate of enzyme catalysed reactions is influenced by changing the environmental conditions. The important factors that influence the velocity of enzyme catalysed reactions are temperature, pH, substrate concentration, and modulators.

1. Temperature

The rate of an enzyme catalysed reaction increases with the increase in temperature up to a maximum and then falls. When a graph is plotted between temperature versus enzyme activity, a bell-shaped curve is obtained (Fig. 4.3). The temperature at which the maximum rate of reaction occurs is called the enzyme's optimum temperature. The optimum temperature is different for different enzymes; but for most of the

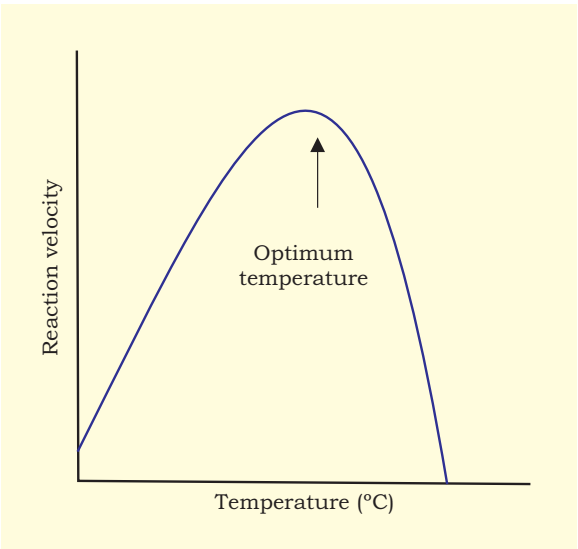


Fig. 4.3: Effect of temperature on enzyme activity

© NCERT not to be republished

enzymes it is between 40-45°C. Majority of enzymes in the human body have an optimum temperature of around 37°C (98.6°F) and are denatured or degraded at extreme temperatures. However, few enzymes like Taq DNA polymerase present in thermophilic bacteria, *Thermus aquaticus,* venom phosphokinase and muscle adenylate kinase are active even at 100°C.

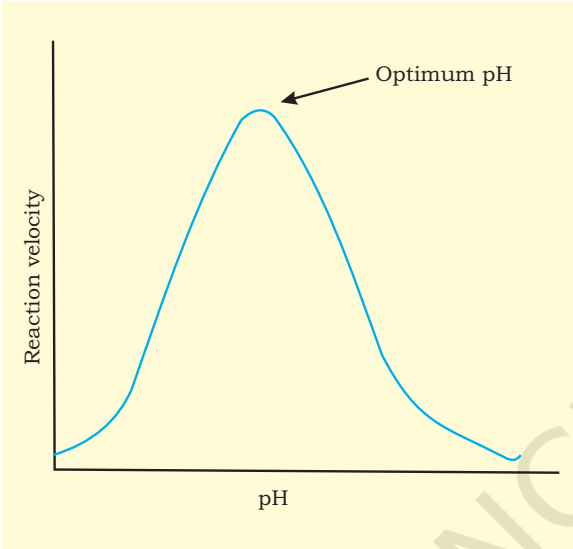


Fig. 4.4: Effect of pH on enzyme activity

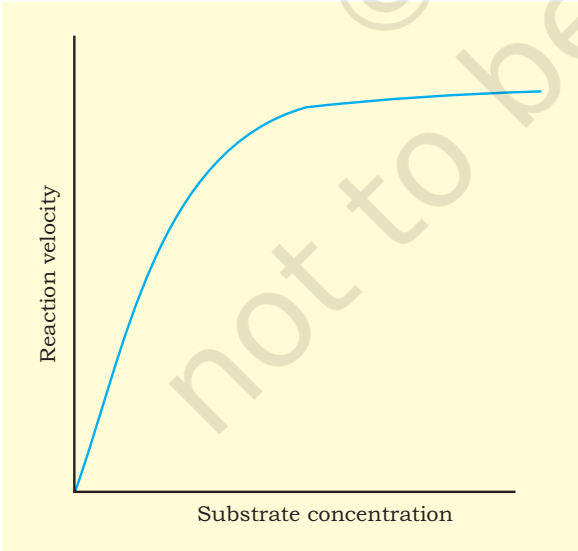


Fig. 4.5: Effect of substrate concentration on reaction rate

2. Hydrogen Ion Concentration (pH)

Enzyme activity is also affected by pH. A plot of enzyme activity against pH results in a bell shaped curve (Fig. 4.4). Each enzyme has its unique optimum pH at which the rate of reaction is greatest. The optimum pH is the pH at which the activity of a particular enzyme is at maximum. Many enzymes of higher organisms show optimum reaction rate around neutral pH (pH 6-8). However, there are several exceptions such as pepsin (pH 1-2), acid phosphatases (pH 4-5) and alkaline phosphatases (pH 10-11). Below and above the optimum pH, the enzyme activity is much lowered and at extreme pH, the enzyme becomes totally inactive.

3. Substrate concentration

The substrate concentration also influences enzyme activity. As the substrate concentration increases the rate of reaction also increases. This is because the more substrate molecules will interact with enzyme molecules, the more products will be formed. However, after a certain concentration, further increase in substrate concentration will have no effect on the rate of reaction, since the substrate concentration will no longer be the limiting factor (Fig. 4.5). At this stage, enzyme molecules become saturated and work at their maximum possible rate.

© NCERT not to be republished

4.1.3 Unit of enzyme activity

The **enzyme unit (U)** is the amount of enzyme that catalyses the conversion of 1 micromole of substrate into product per minute under standard conditions. The International Union of Biochemistry (I.U.B.) adopted enzyme unit as unit of enzyme activity in 1964. But it was discouraged in favour of the **katal** since the minute is not an SI unit. One katal (kat) is the amount of enzyme that catalyses conversion of 1 mole of substrate per second into product (1 kat = 6×10^7 IU).

4.1.4 Specific activity

Specific activity is the unit of enzyme activity per milligram of enzyme protein. Specific activity represents a measurement of enzyme purity in the mixture.

4.1.5 Mechanism of enzyme action

For understanding the enzyme mechanism, you should consider two thermodynamic properties of a reaction. These are the free energy difference (ΔG) between products and reactants and the energy needed to initiate the conversion of reactant into product. The former energy, i.e., ΔG determines whether the reaction is spontaneous, whereas the latter determines the rate of reaction. Enzymes affect the energy, which determines the rate of reaction. The enzymes cannot change the laws of thermodynamics and therefore, cannot alter the equilibrium of a biochemical reaction. They speed up the attainment of equilibrium.

Rate of reaction rather depends on the free energy of activation (ΔG^A), which is not related to ΔG. The substrate S of a reaction converted into product P via formation of a transition state which has higher free energy than either S or P. The difference between the free energy of transition state and substrate is called **Gibbs free energy of activation** or simply **activation energy (ΔG^A)**. The enzymes enhance reaction rate without altering ΔG of the reaction, rather they lower the activation energy, ΔG^A.

Kinetics of enzyme-catalysed reaction

During catalysis, substrate S binds at the active site of the enzyme E and results in formation of enzyme substrate complex ES, which is finally converted into product P. The reaction can be represented as:

© NCERT not to be republished

$$E+S \rightleftharpoons [ES] \rightarrow E+P$$

Where, E forms weakly bonded complex ES with the substrate S. The ES complex decomposes to yield product P and the free enzyme E.

The kinetics of enzyme catalysed reactions was explained by Leonor Michaelis and Maud Menten in 1913. The most remarkable feature of this kinetics is that specific ES complex is an intermediate during catalysis. Michaelis-Menten theory of enzyme kinetics is the simplest one that accounts for kinetic properties of many enzymes.

Further simplifying the reaction, Michaelis-Menten derived following equation for one substrate reaction.

$$v_0 = \frac{V_{max}\ [S]}{K_m\ + [S]}$$

The equation is called Michaelis-Menten equation. Where, K_m is called Michaelis constant, v_0 is initial velocity, V_{max} is maximum velocity of reaction, and [S] is substrate concentration.

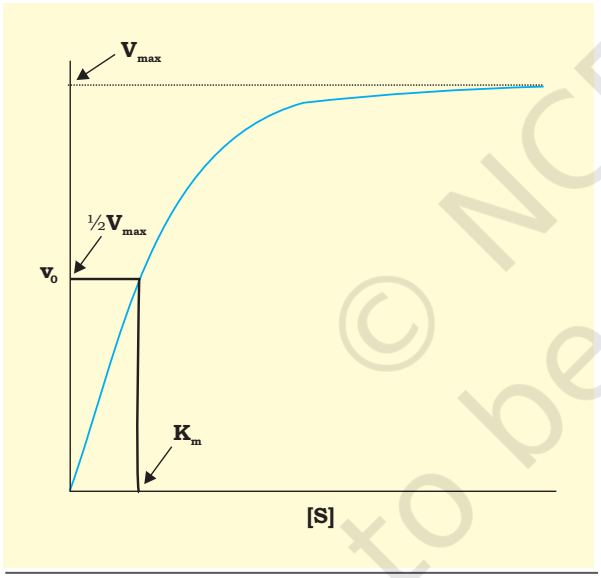


Fig. 4.6: Graph of v_0 against [S] at constant enzyme concentration for a single substrate enzyme-catalysed reaction for Michaelis-Menten equation

A graph of v_0 against [S] results in rectangular hyperbola (Fig. 4.6). V_{max} is the maximum velocity at particular enzyme concentration. V_{max} and K_m can be determined from the graph as shown in Fig. 4.6.

In the graph, we can see that at very low substrate concentration (when $[S] << K_m$), $v_0=(V_{max}/K_m)/[S]$, i.e., the reaction rate is directly proportional to substrate concentration. At high substrate concentration (when $[S] >> K_m$), $v_0=V_{max}$, i.e., reaction rate is maximum and independent of substrate concentration. When $[S]=K_m$, then $v_0= V_{max}/2$. Thus, K_m is substrate concentration at which half of the maximum reaction rate is obtained.

4.1.6 Enzyme inhibition

Substances which decrease the rate of an enzyme catalysed reaction are called as **enzyme inhibitors** and the process is known as enzyme inhibition.

© NCERT not to be republished

Enzyme inhibition can be classified as **reversible inhibition** and **irreversible inhibition**. In irreversible inhibition, the inhibitor binds very tightly to the enzyme and does not dissociate from it. For example, the antibiotic penicillin acts as inhibitor and binds with the enzyme transpeptidase, which is responsible for synthesis of bacterial cell wall. Hence, binding of this drug to the enzyme prevents cell wall synthesis, thus killing the bacteria. In the same way, the drug aspirin inhibits the enzyme cyclooxygenase, thus reducing the inflammation.

In reversible inhibition, the inhibitor rapidly dissociates from the enzyme-inhibitor complex. There are three types of reversible inhibitions: **competitive, non-competitive** and **uncompetitive inhibition**.

(i) Competitive inhibition

In competitive inhibition, there is close resemblance in the structure of inhibitor I and substrate S, therefore, they both compete for the same active site on the enzyme. The enzyme can form enzyme-substrate ES complex or it can form enzyme-inhibitor EI complex (Fig. 4.7) but not ESI.

Competitive inhibitors decrease the rate of reaction by reducing the amount of active enzyme molecules bound to a substrate. At very high substrate concentration, the chances of binding of inhibitor molecule to the enzyme will be reduced, so V_{max} for the reaction will not be changed. However, the K_m is increased in presence of inhibitor and is denoted by symbol K'_m (Fig. 4.8).

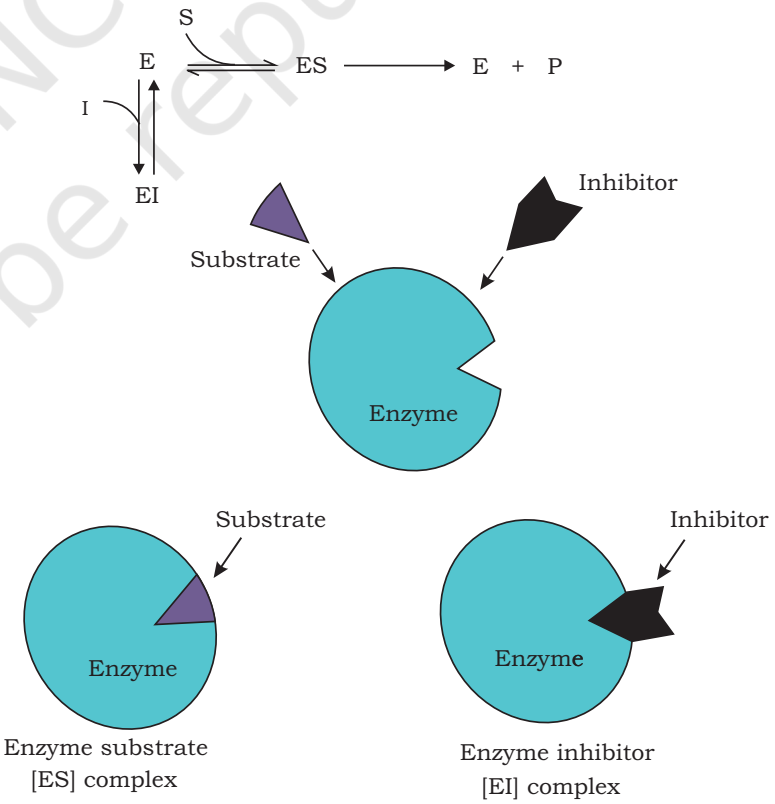


Fig. 4.7: Competitive inhibition

(ii) Non-Competitive inhibition

In this type of inhibition, inhibitor has no structural similarity with substrate and binds with the enzyme at different site other than the active site. Therefore, the

© NCERT not to be republished

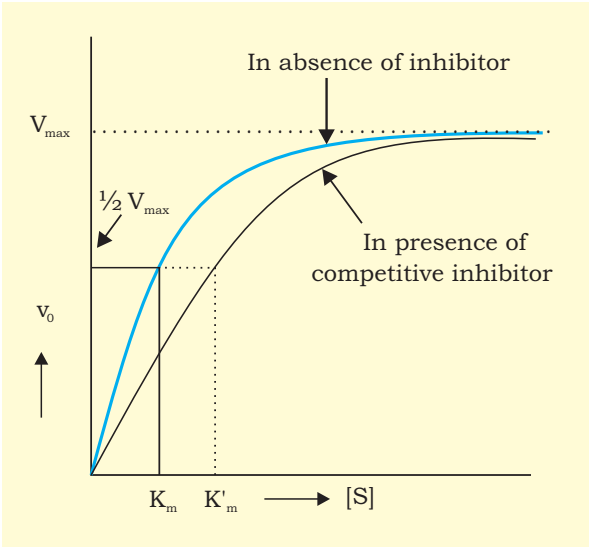


Fig. 4.8: Michaelis-Menten plot for competitive inhibition

inhibitor can bind with E as well as ES complex and formation of ES, EI and ESI takes place (Fig. 4.9).

The inhibitor I and substrate S can bind simultaneously to the same enzyme molecule as their binding sites are different and hence do not overlap (Fig. 4.9). Non-competitive inhibitor lowers the V_{max} rather than by decreasing the proportion of enzyme molecules that are bound to the S. Thus, the non-competitive inhibition in contrast to competitive inhibition, cannot be overcome by increasing substrate concentration. The substrate can still bind to the EI complex. However, the ESI does not form product. The I effectively lowers the concentration of active enzyme

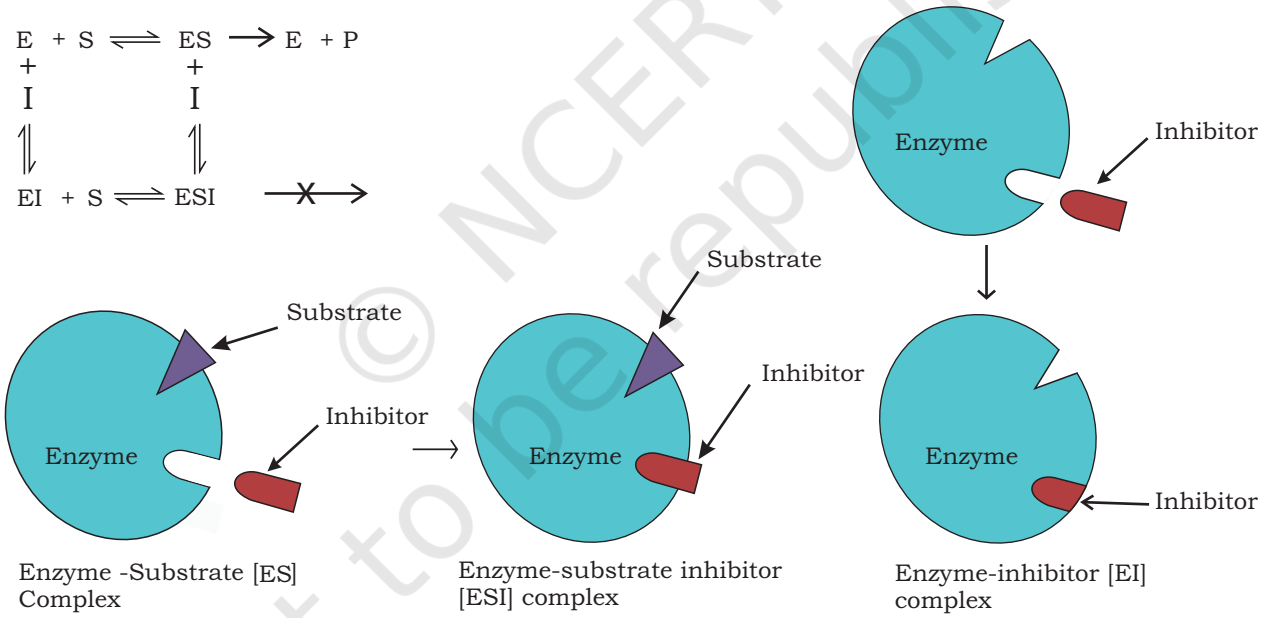


Fig. 4.9: Non-competitive inhibition

and hence lowers the V_{max}. There is no effect on K_m as the inhibitor decrease the amount of functional enzyme (Fig. 4.10).

(iii) Uncompetitive inhibition

In this type of inhibition, inhibitor does not bind to free enzyme. It binds only to enzyme substrate (ES) complex (Fig. 4.11). The inhibitor does not compete with substrate for the same binding site. Therefore, the inhibition cannot be overcome by increasing substrate concentration.

© NCERT not to be republished

Both K_m and V_{max} values are altered.

4.1.7 Allosteric enzymes

Allosteric enzymes do not obey Michaelis-Menten kinetics. These enzymes result in sigmoidal graph instead of rectangular hyperbola when v_0 is plotted against substrate concentration [S] (Fig. 4.12). Each subunit of allosteric enzymes also contain regulatory site along with active site. Regulatory molecules may reversibly bind to the regulatory site and alter the affinity of enzyme for substrate binding. Whereas, most of the enzymes obeying Michaelis-Menten kinetics are the normal enzymes, allosteric enzymes are the key regulators of metabolic pathways in the cell.

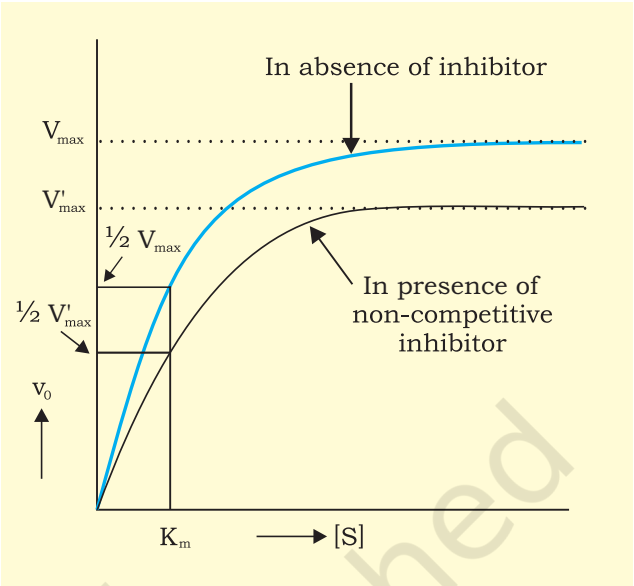


Fig. 4.10: Michaelis Menten plot for non-competitive inhibition

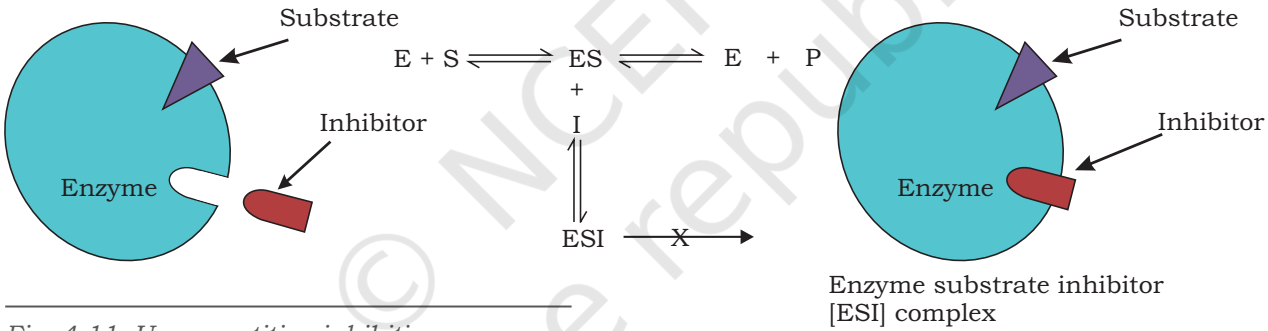


Fig. 4.11: Uncompetitive inhibition

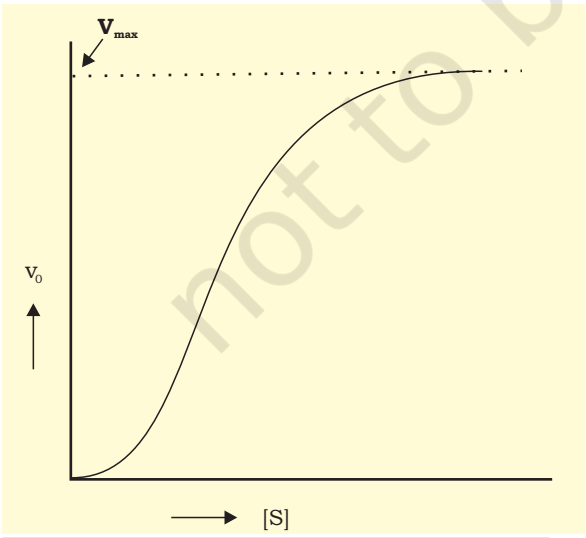


Fig. 4.12: Kinetics of an allosteric enzyme

© NCERT not to be republished

4.2 BRIEF INTRODUCTION TO BIOENERGETICS

Biological energetics or bioenergetics deal with the transformation and use of energy by living cells. In biological reactions, the energy is released as the reactions move from a higher to lower energy level. Energy liberated is utilised to perform work. The energetics of these processes are governed by the laws of thermodynamics, as described below.

4.2.1 The laws of thermodynamics

Thermodynamics is the branch of physical chemistry that deals with the energy changes. Energy is the capacity to do work and it exists in a variety of forms such as electrical, mechanical, chemical, heat and light. These forms of energy are interconvertible. Thus, bioenergetics is concerned with the changes in energy during biochemical processes. It does not deal with the mechanism or speed of the process. There are two fundamental laws of thermodynamics which explain interconversions of various forms of energy. The two laws also help to understand the following:

1. The direction of a reaction, whether forward or reverse.
2. The accomplishment of work, whether useful or non-useful.
3. Whether the energy for executing a reaction must be delivered from an external source.

The first law of thermodynamics

The first law of thermodynamics states that 'energy can neither be created nor destroyed, but can be converted into other forms of energy'. It means, the total amount of energy in the universe (system plus surroundings) remains constant. According to this law, energy exchange in any process takes place between 'system' and 'surroundings'. A system is a matter within defined region, and surroundings constitute matter in the rest of the universe. Thus, the system and surroundings make the universe, which includes the entire earth and the outer space. Thus, the first law of thermodynamics can be expressed in the form of following equation:

$$\Delta E = E_B - E_A = Q - W$$

Where,

ΔE = change in internal energy

© NCERT not to be republished

E_A = energy of a system at the start of a process

E_B = energy of a system at the end of a process

Q = heat absorbed by the system

W = work done by the system

According to this equation, change in the energy of a system depends only on the initial and the final stages, and not the path of the transformation.

The second law of thermodynamics

The first law of thermodynamics cannot be used to predict the spontaneity or non-spontaneity of the reaction. On the other hand, the second law introduces a term '**entropy**', which is denoted by symbol 'S'. Entropy is degree of randomness or disorder of a system and explains whether a reaction takes place spontaneously. The entropy of a system increases when it becomes more disordered. Entropy reaches maximum in a system as it nears equilibrium. Biological systems, however, do not follow the second law as the life is state of higher organisation or lower entropy rather than increase entropy. This highly ordered form of life is maintained for a while by consumption of chemical energy (food) and in case of photosynthetic organisms through light energy. This energy is either converted to a less organised form of energy (heat) or utilised to perform work. Ultimately, the thermodynamic equilibrium is certain and entropy increases after death and subsequent decomposition of every organism.

The second law of thermodynamics states that the entropy or disorder of universe is always increasing. According to this law, a process can occur spontaneously only if the sum of the entropies of the system and its surrounding increases. This can be represented as:

$(S_{system} + S_{surroundings}) > 0$ (for a spontaneous process)

Thus, for a process to occur spontaneously, the total entropy of a system must increase. However, the entropy of a system can decrease even during a spontaneous process, provided the entropy of the surrounding increases to an extent that their sum becomes positive.

The difference between the first and second law is that, first law is concerned with the transformation of various kinds of energy involved in a given process, whereas the

© NCERT not to be republished

second law is concerned with the availability of the energy of a given system for doing work.

Combining the two laws

Entropy is not used as a criteria to know whether a biochemical reaction is spontaneous or not. This is because changes in entropy of reaction cannot be measured. Moreover, changes in entropy of surrounding and the system should be known for spontaneity. Therefore, a different thermodynamic function '**free energy**' is used to overcome this problem. Free energy is denoted by symbol ΔG.

In 1878, **Gibbs** created free energy function by combining the first and second laws of thermodynamics. Following equation was obtained,

$$\Delta G = \Delta H - T\Delta S$$

Where,

ΔG = change in free energy of a reacting system

ΔH = change in heat content or enthalpy of this system

T = absolute temperature at which the process is taking place

ΔS = change in entropy of the system

This equation shows the relationship between the change in free energy (ΔG), heat (also called **enthalpy**, ΔH), and entropy in chemical reactions at constant temperature (T) and pressure (P). Biochemical reactions also occur at these conditions. The ΔG is free energy change or theoretically available useful work. The term TΔS is that component of ΔH which cannot be used to perform work.

A closed system is described as system that can exchange energy but not matter with its surroundings. The exchange of energy must involve heat transfer or the performance of work. If, in a closed system at constant temperature and pressure, a process takes place which involves a transfer of heat to or from the surroundings and a change in volume of the system, then from the first law of thermodynamics,

$$\Delta E = \Delta H - P\Delta V$$

Where,

ΔE = change in internal energy

ΔH = change in enthalpy

© NCERT not to be republished

PΔV = work done on the surroundings by increasing the volume of the system by ΔV at constant pressure P and temperature T.

ATP: the universal currency of free energy

The living organisms derive free energy from the environment. The photosynthetic organisms take this energy from sunlight whereas, chemotrophs (non photosynthetic organisms) obtain it by oxidation of food stuff. This free energy is used to complete some vital processes in the cell such as (i) synthesis of macromolecules from small precursors, (ii) in active transport (transport against gradient) across the membrane, (iii) in muscle contraction, and (iv) in the fidelity of genetic information transfer. Before being utilised in the above-mentioned process the free energy (derived from light or from the oxidation of food stuff) is partly converted into a special form, **adenosine triphosphate (ATP)**. ATP is also known as universal currency of free energy. It plays a central role in the transfer of free energy from **exergonic** (energy releasing) processes to **endergonic** (energy consuming) processes in the cells. During the breakdown of the energy-rich food

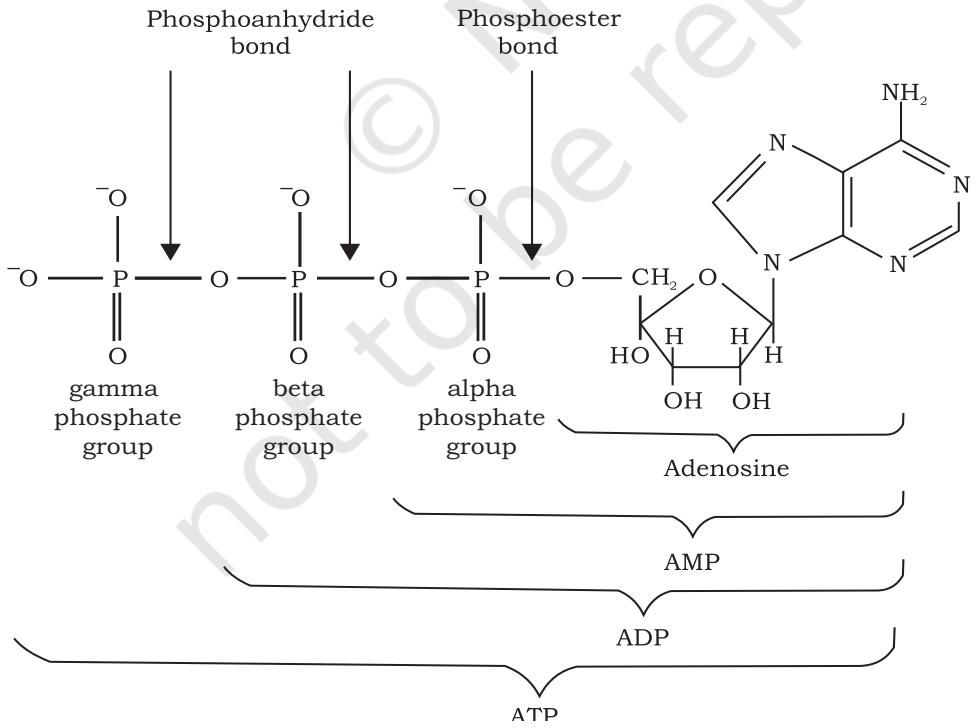


Fig. 4.13: Structure of AMP, ADP, and ATP

© NCERT not to be republished

stuff, some of the free energy is consumed in synthesis of ATP from **adenosine diphosphate (ADP)** and inorganic phosphate (Pi). ATP then donates much of its chemical energy to energy-requiring processes such as biosynthesis, transport, muscle contraction, etc., by converting itself into ADP and Pi and releases 7.3 Kcal/mol of energy. ATP can be converted into **adenosine monophosphate (AMP)** and PPi (pyrophosphate) while donating its energy in some other processes e.g., in the production of light flashes by firefly. ATP and its consecutive hydrolysis products, ADP and AMP are nucleotides, made up of an adenine (a purine base occur in DNA and RNA) a ribose (a pentose sugar) and three, two and one phosphate group(s), respectively (Fig. 4.13).

SUMMARY

- Enzymes are catalysts that catalyse the biochemical reactions in the living system.
- Each enzyme has an active site into which the substrate molecule fits precisely. The 'lock and key' hypothesis postulated that the substrate fits precisely into the lock of the enzyme. This hypothesis has now been modified.
- The modern 'induced fit' hypothesis does not regard active site as a rigid structure but a flexible one, which modifies its shape to fit precisely the substrate molecule.
- Various factors like temperature, pH, substrate concentration and presence of inhibitors and activators influence the rate of enzyme catalysed reactions.
- Enzymes lower the activation energy of the reaction they catalyse.
- Simple single substrate enzyme catalysed reactions can be described by Michaelis-Menten kinetics which has a hyperbolic graph in terms of substrate concentration and initial velocity.
- Enzymes are also affected by the presence of inhibitors, like competitive, non-competitive, and uncompetitive inhibitors, which slow down the rate of reaction or stop it completely.

© NCERT not to be republished

- Bioenergetics deals with the transformation and use of energy by living cell.
- There are two laws of thermodynamics. The first law states that energy can neither be created nor destroyed, but can be converted into other forms of energy.
- The second law of thermodynamics states that entropy or disorder of universe is always increasing.
- Combining the two laws of thermodynamics describe the free energy (G) change or the performance of work.
- The photosynthetic organisms derive free energy from the sunlight, whereas chemotrophs obtain free energy by oxidation of food stuff.
- This free energy is used to complete the cellular processes.
- This free energy is partly converted into ATP which is also known as universal currency of free energy.
- During energy releasing processes (breakdown), ATP is synthesised from ADP and inorganic phosphate (Pi), while in energy requiring processes ATP is broken into ADP and Pi.

EXERCISES

1. In order to catalyse a reaction, an enzyme is required to
   (a) be saturated with substrate
   (b) decrease the activation energy
   (c) increase the equilibrium constant
   (d) increase the activation energy
2. Pepsin is a gastric enzyme. Does it have an acidic or alkaline optimum pH? What happens to pepsin when it enters the duodenum?
3. What is the relationship between vitamins and enzyme co-factors?
4. What is the effect of temperature, pH, and substrate concentration on catalytic activity of enzyme?
5. The rate determining step of Michaelis-Menten kinetics is
   (a) the complex dissociation of ES complex
   (b) the complex formation
   (c) the product formation
   (d) the product degradation

© NCERT not to be republished

6. Define K_m and its significance.
7. What is meant by one unit of enzyme?
8. What is specific activity of an enzyme?
9. Briefly describe first and second laws of thermodynamics.
10. Define entropy. What is the relationship between free energy and entropy?
11. Why ATP is called universal energy currency?

© NCERT not to be republished

11150CH05

Chapter 5

Cellular Processes

5.1 *Cell Signaling*
5.2 *Metabolic Pathways*
5.3 *Cell Cycle*
5.4 *Programmed Cell Death (Apoptosis)*
5.5 *Cell Differentiation*
5.6 *Cell Migration*

5.1 Cell Signaling

We are aware that cells (both prokaryotic and eukaryotic) constantly receive and interpret environmental signals and respond to them in real time. These signals include light, heat, sound, and touch. The cell fates during development are specified by signaling pathways in response to such extracellular signals. Cells interact with their neighbouring cells by transmitting and receiving signals. These signals are synthesised by the cells in the form of chemicals and released in the extracellular milieu. However, cells can also respond to 'external' signals which are not synthesised by the cells of our body. Therefore, one can assume that the cells are capable of sensing a wide variety of signals. It is important to note that a cell can only respond to a particular signal if it possesses the corresponding receptor for it. A **receptor** is a protein located either on the cell surface or inside the cytoplasm or the nucleus. A chemical messenger to which a receptor responds is a **ligand**. The association between a receptor and its corresponding ligand is highly specific, which means that a cell will only

© NCERT not to be republished

be able to respond to a chemical messenger, if it bears the corresponding receptor for it and not otherwise.

Transmission of chemical messages from one cell to another cell requires binding of a ligand to its receptor, which results in conformational changes in the receptor. These changes then initiate a message relay system and bring about further important changes in the activities inside the cell.

It should be noted that cells send and receive signals in different ways. Depending on the proximity of sender and recipient cells, signaling can be broadly classified in the following categories:

1. **Paracrine signaling:** In this form of signaling, communication between cells occurs over relatively short distances. A chemical message released in the extracellular space by the sender cells is sensed by the recipient cells instantly. This type of signaling is seen commonly in the communication of neurons.
2. **Autocrine signaling:** Many times, a cell which secretes a ligand, also possesses receptors specific for that ligand. This type of signaling is referred to as autocrine signalling. For instance, cancer cells are characterised by uncontrollable growth. Therefore, they require a greater amount of growth factors for their proliferation. Unlike normal cells, cancer cells do not depend on external growth factors for their growth. Instead, they are capable of synthesising their own growth factors and also possess the receptors specific for them.
3. **Endocrine signaling:** Endocrine signaling or long-distance signaling requires the ligand to be synthesised by the cell and released into the bloodstream to travel to the recipient or target cell. Hormones generally exhibit this form of signaling.

5.2 METABOLIC PATHWAYS

Metabolism is the process through which living organisms take and utilise the free energy required to carry out their life processes. Living organisms are of two types on the basis of taking free energy: phototrophs and chemotrophs.

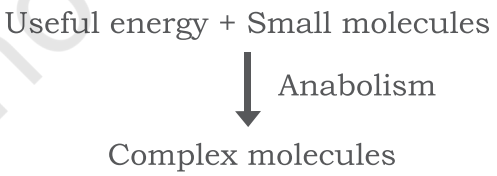
© NCERT not to be republished

Phototrophs use the energy of sunlight to convert simple molecules (less energy containing) into more complex molecules (energy rich) that serve as fuel to perform life processes. Phototrophs are photosynthetic organisms (such as plants and some bacteria); they transform light energy into chemical energy. **Heterotrophs** such as animals, obtain energy indirectly from plants through their food. In **chemotrophs,** the energy is obtained by oxidising chemical compounds (organic or inorganic). This energy uptake in organisms is done by coupling the exergonic reactions of nutrient oxidation to the endergonic processes required to maintain the living state. Central to all these energy transactions is the energy currency called ATP (detail is given in section 4.2 bioenergetics). In metabolism, there are interlinked biochemical reactions that begin with a particular molecule and convert it into some other molecule or molecules in a carefully defined fashion. The energy is utilised for various processes within the cell such as, the creation of gradient, movement of molecules across membranes, conversion of chemical energy into mechanical energy and powering of reactions that result in the synthesis of biomolecules.

The synthesis and breakdown of biomolecules is accomplished through a number of steps inside the living system. These steps collectively constitute metabolic pathway. Metabolic pathways can broadly be classified into two classes; anabolic pathways and catabolic pathways.

(i) Anabolic pathways

In these pathways, larger and more complex molecules are synthesised from small molecules. Anabolic pathways are endergonic (consumption of energy). Reactions that require energy such as synthesis of glucose, fats, protein or DNA are called anabolic reactions or anabolism.



© NCERT not to be republished

(ii) Catabolic pathways

These pathways involve the breakdown of larger molecules. These are exergonic (release of energy) reactions and produce reducing equivalents and ATP. The useful forms of energy that are produced in catabolism are utilised in anabolism, to generate complex structures from simple ones or energy-rich states from energy poor ones.

Fuel (carbohydrate, protein, fats)

↓ Catabolism

CO_2 + H_2O + Useful energy

5.2.1 Overview of carbohydrate metabolism

In animals, the metabolic fuel for most of the tissues is glucose. Glucose is metabolised into pyruvate through glycolysis. In aerobic condition (in presence of oxygen) pyruvate enters into mitochondrial matrix, where it is converted into acetyl CoA and take part in the citric acid cycle to complete oxidation of glucose to CO_2 and H_2O (Fig. 5.1). This oxidation is linked to the formation of ATP through the process of oxidative phosphorylation. In anaerobic (in absence/lack of O_2) condition pyruvate is converted into lactic acid. The metabolic intermediates of glycolysis also take part in other metabolic processes, such as

(i) In synthesis of glycogen and its storage in animals.

(ii) In pentose phosphate pathway which is source of reducing equivalent (NADPH) for fatty acid synthesis, and source of ribose for nucleotides and nucleic acid synthesis.

(iii) The triose phosphate generates glycerol moiety of triacylglycerol.

(iv) Acetyl CoA is the precursor for synthesis of fatty acids and cholesterol. Cholesterol then synthesises all other steroids in animals.

(v) Pyruvate and intermediates of citric acid cycle give rise to carbon skeleton for amino acid synthesis.

(vi) When glycogen reserves are depleted such as in starvation conditions the non-carbohydrate precursors such as lactic acid, amino acids, and glycerol can

© NCERT not to be republished

synthesise glucose through the process of gluconeogenesis.

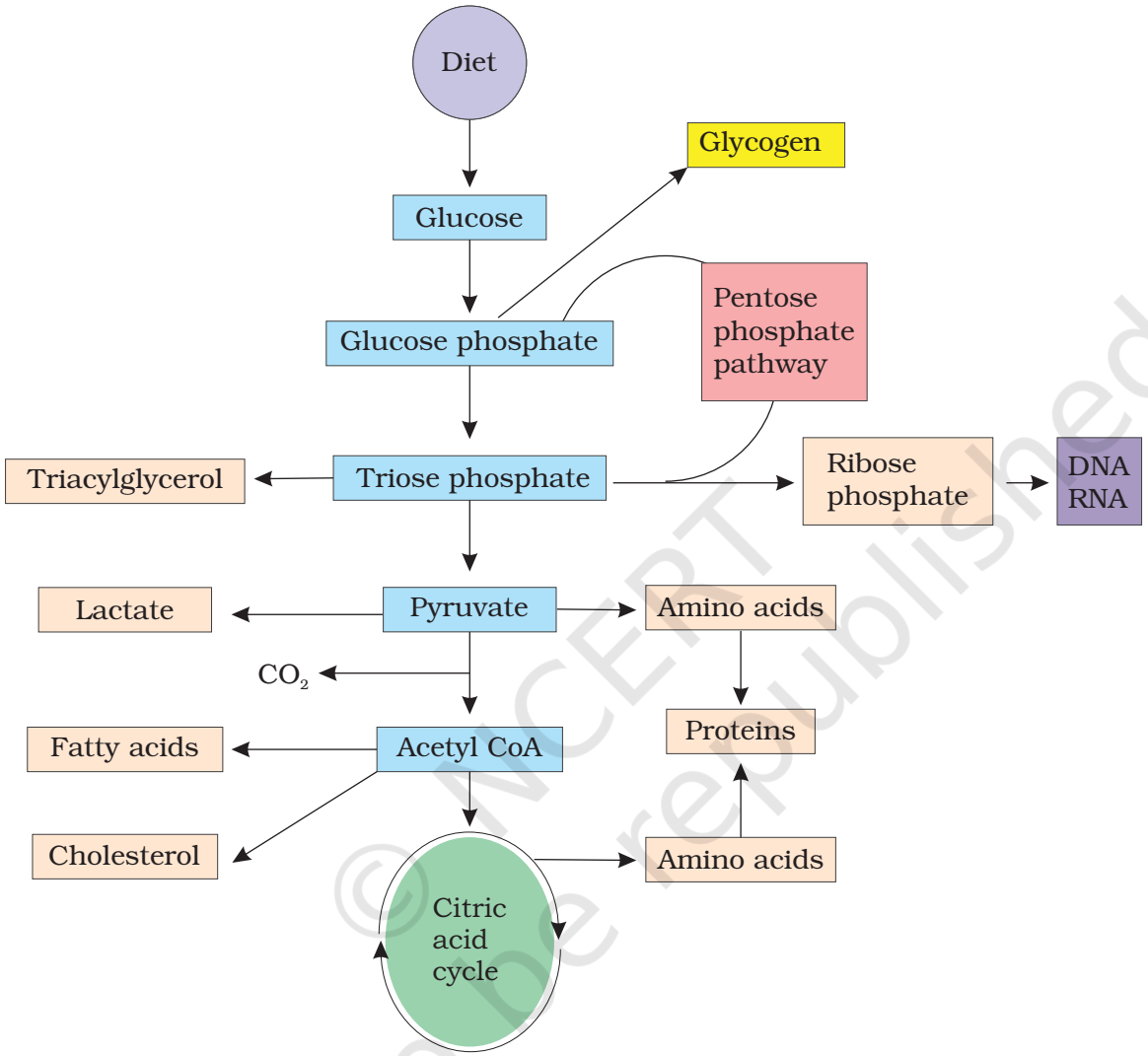


Fig. 5.1: Overview of carbohydrate metabolism

5.2.2 Overview of lipid metabolism

Some vital tissues such as brain, heart and red blood cells are exclusively dependent on glucose. In the fasting state when glucose is limiting, then less glucose-dependent tissues such as muscles, liver and other tissues alternatively use fuel other than glucose (Fig. 5.2). This fuel is long chain fatty acids which are either taken from diet or synthesized from acetyl CoA derived from carbohydrate or amino acids. Fatty acids may be oxidized to acetyl CoA through the β-oxidation pathway or esterified with glycerol

© NCERT not to be republished

to synthesize triacylglycerol (fat) as main fuel reserve in adipose tissue of animals. Following are three fates of acetyl CoA formed by the β-oxidation pathway.

(i) It is oxidized to CO_2 and H_2O through the citric acid cycle.

(ii) It is a precursor for the synthesis of other lipids such as cholesterol. Cholesterol then synthesises all other steroids (hormones and bile pigments).

(iii) It is used to synthesise ketone bodies (acetone, acetoacetate and 3-hydroxy butyrate) which are an alternative fuel for liver, and some other tissues in prolonged fasting.

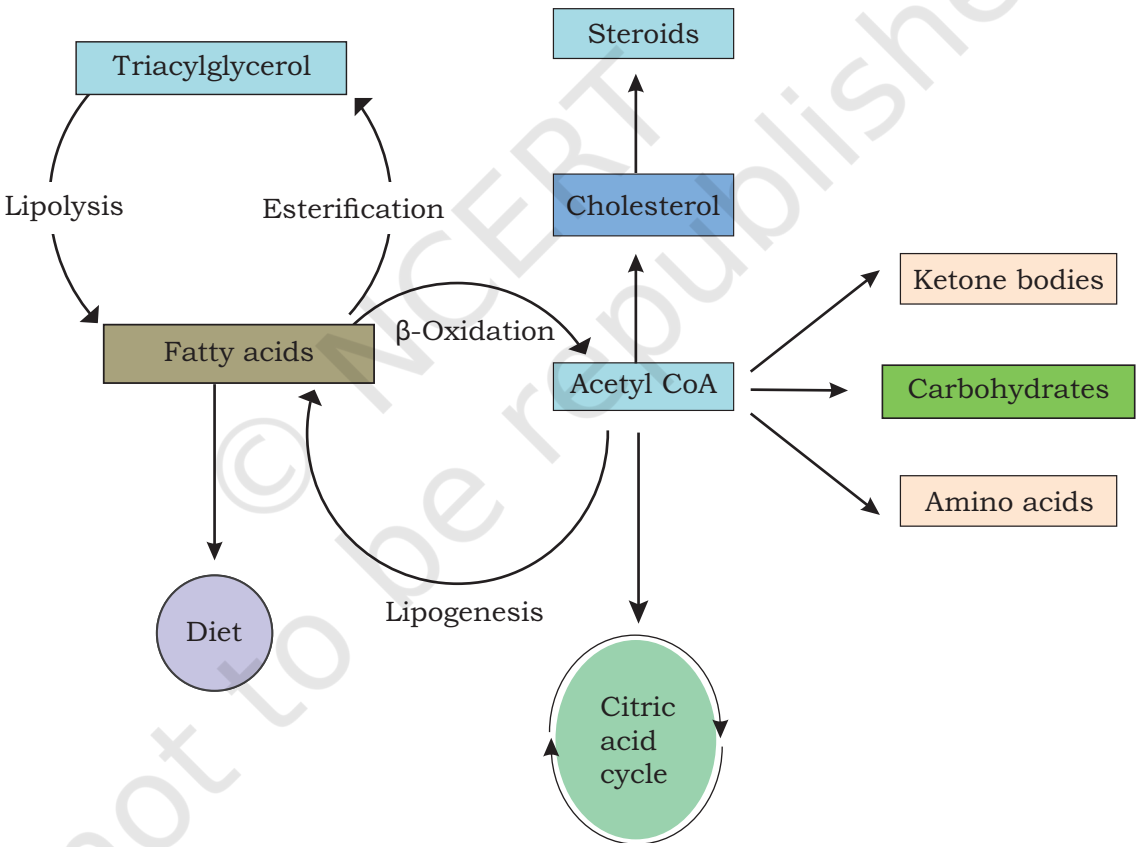


Fig. 5.2: Overview of lipid metabolism

5.2.3 Overview of amino acid metabolism

Since amino acids are building blocks of proteins, therefore they are required for protein synthesis. There are 20

© NCERT not to be republished

standard amino acids. Some are non essential amino acids as these are synthesised in the body through metabolic intermediates by the process of transamination (Fig. 5.3). The remaining are essential amino acids which must be supplied in the diet as they are not synthesised in the body. In transamination, the amino nitrogen is transferred from one amino acid to a carbon skeleton to form other amino acids. In the process of deamination, the amino nitrogen is excreted as urea. The carbon skeletons that remain after transamination can play the following roles:

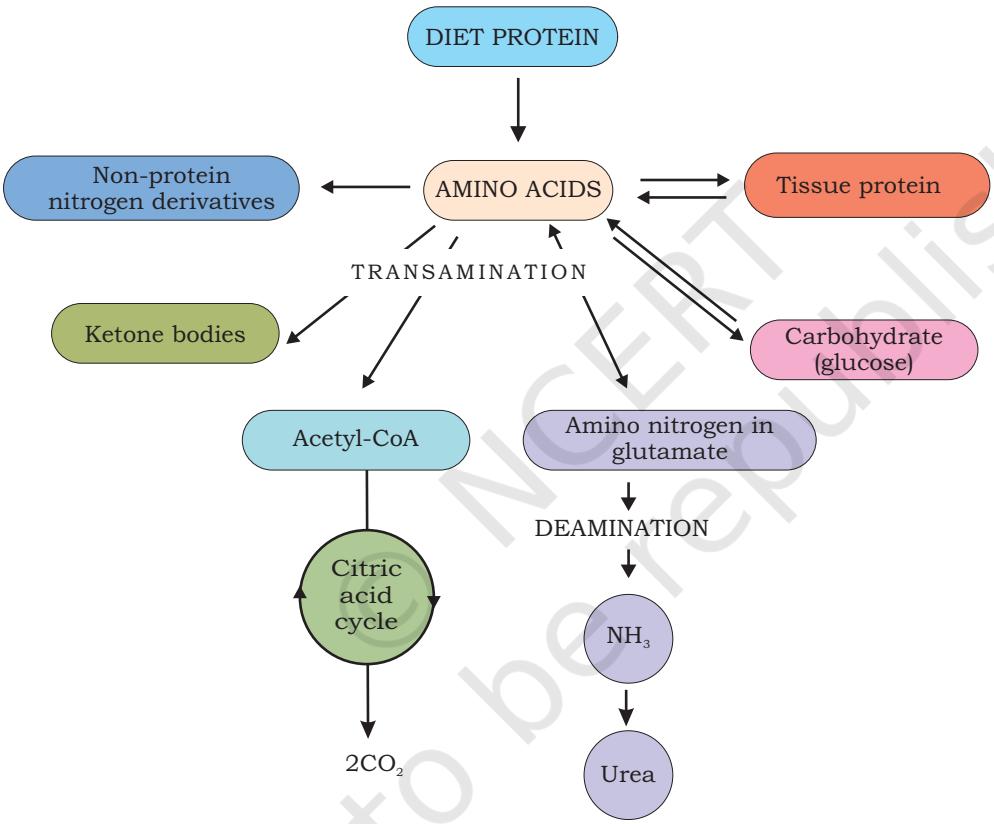


Fig. 5.3: Overview of amino acid metabolism

(i) Oxidised to CO_2 via citric acid cycle.

(ii) Used to synthesise glucose through gluconeogenesis.

(iii) Form ketone bodies, which may be oxidised or used for fatty acid synthesis.

Some amino acids take part in the synthesis of other biomolecules like hormones of plant and animals, purines, pyrimidines, and neurotransmitters.

© NCERT not to be republished

Some of the important metabolic pathways are described and detailed in the following section.

5.2.4 Glycolysis

Glycolysis is a universal catalytic pathway in all living cells, also known as Embden-Meyerhof-Parnas (EMP) pathway (Fig. 5.4). It is a major pathway of carbohydrate metabolism and all the enzymes involved are present in the cytosol, and the pathway starts with the phosphorylation of glucose (6 carbon molecule) to glucose-6-phosphate which is catalysed by enzyme hexokinase. ATP is phosphate donor; its γ-phosphoryl group is transferred to glucose. This reaction is irreversible and enzyme hexokinase is inhibited allosterically (when product binds to the enzyme at a site different from the active site and alters its catalytic activity) by its product glucose-6-phosphate. Hexokinase can also phosphorylate sugar other than glucose such as fructose, galactose, mannose, etc. Liver cells also contain an isoenzyme (detail is given in section 4.1) of hexokinase called glucokinase which can phosphorylate only glucose. Glucose-6-phosphate is an important intermediate of carbohydrate metabolism as it is formed in glycolysis, gluconeogenesis (formation of glucose from noncarbohydrate molecules), pentose phosphate pathway, glycogenesis (synthesis of glycogen) and glycogenolysis (breakdown of glycogen).

Glucose-6-phosphate is then converted into fructose-6-phosphate (6 carbon molecule) by enzyme phosphoglucose isomerase which catalyses an aldose-ketose isomerisation reaction. Fructose-6-phosphate then undergoes another phosphorylation by the enzyme phosphofructokinase (PFK) forming fructose-1-6-bisphosphate. Similar to hexokinase, PFK also catalyzes the irreversible reaction and undergoes allosteric regulation. Fructose-1-6-bisphosphate is cleaved by enzyme aldolase into two triose phosphates (each 3 carbon molecule), glyceraldehyde-3-phosphate and dihydroxyacetone phosphate.

The two trioses synthesised are interconverted by the enzyme triose phosphate isomerase. Oxidation of glyceraldehyde-3-phosphate to 1-3-bisphosphoglycerate occur by the enzyme glyceraldehyde-3-phosphate dehydrogenase, which is a NAD-dependent enzyme. In the next reaction, phosphate is transferred from 1-3-bisphosphate glycerate to ADP, forming ATP and 3-phosphoglycerate

© NCERT not to be republished

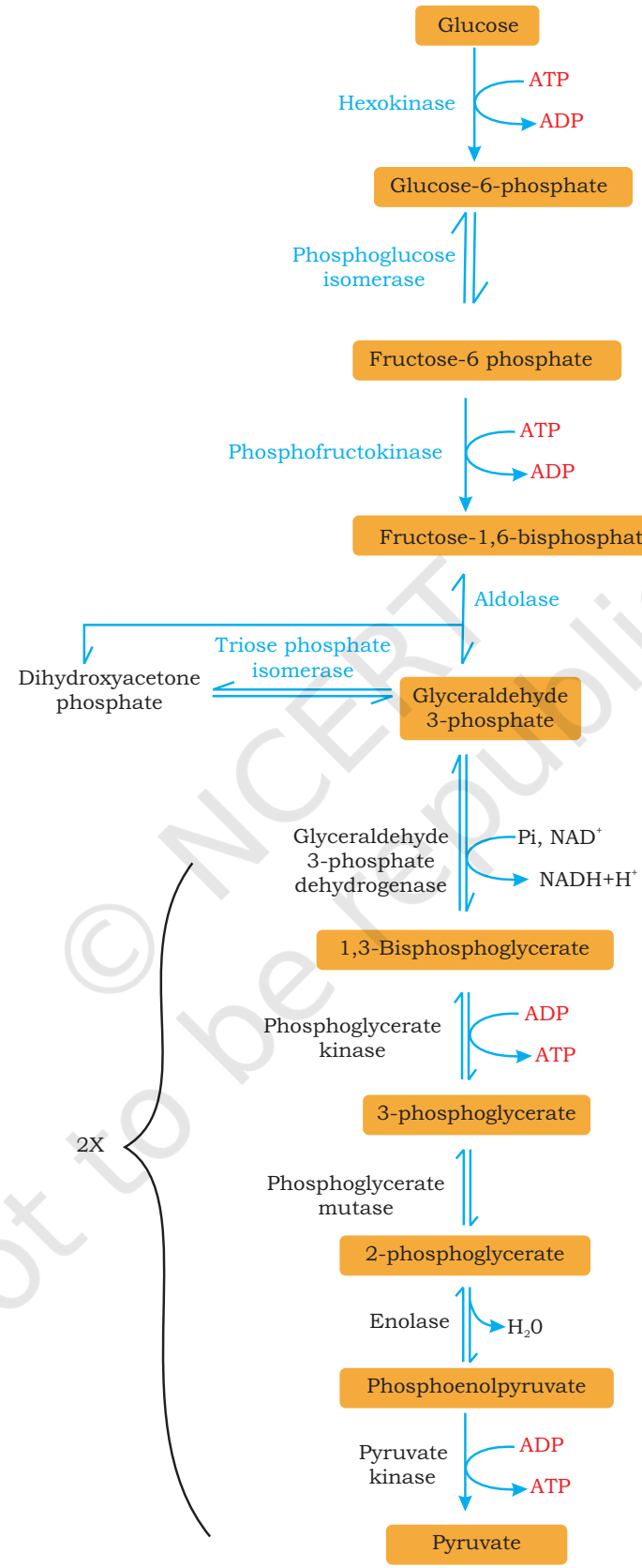


Fig. 5.4: Glycolysis

© NCERT not to be republished

(3 carbon molecule) by the enzyme phosphoglycerate kinase. This phosphorylation of ADP to form ATP is called substrate-level phosphorylation.

Since two molecules of triose phosphates are synthesised per molecule of glucose, two ATP molecules are formed at this stage per molecule of glucose undergoing glycolysis. Next step of glycolysis involves isomerisation of (3 carbon molecule) 3-phosphoglycerate into 2-phosphoglycerate (3 carbon molecule) by enzyme phosphoglycerate mutase.

The subsequent step involves a dehydration reaction converting 2-phosphoglycerate into phosphoenol pyruvate by the enzyme enolase which requires either Mg^{2+} or Mn^{2+} for its activity. The phosphoenol pyruvate, by enzyme pyruvate kinase is converted into pyruvate (3 carbon molecule), during which phosphate is transferred to ADP to form ATP (substrate level phosphorylation).

Out of the 10 glycolytic reactions, three reactions are exergonic and therefore irreversible. These reactions are catalysed by regulatory enzymes namely, hexokinase, PFK and pyruvate kinase and are therefore major sites of regulation of glycolysis.

Energy yield by glycolysis

The net reaction in the transformation of one molecule of glucose into two pyruvate molecules is:

$$\text{Glucose} + 2\text{ATP} + 2\text{NAD}^{+} + 4\text{ADP} + 4\text{P}_i \rightarrow 2\text{Pyruvate} + 2\text{ADP} + 2\text{NADH} + 2\text{H}^{+} + 4\text{ATP} + 2\text{H}_2\text{O}$$

Thus four molecules of ATP are generated in the conversion of glucose into two molecules of pyruvate. The net ATP production is two as two ATP molecules are utilised during the process.

Fate of pyruvate

There are two major fates of pyruvate form during glycolysis. One is anaerobic (in the absence of oxygen) and other is aerobic (in the presence of oxygen). In the anaerobic fate the pyruvate is metabolised to lactate (which is called homolactic fermentation) or it may be metabolised into ethanol (called alcoholic fermentation). Under aerobic condition, pyruvate is metabolised into CO_2 and H_2O through citric acid cycle (TCA cycle). All the steps

© NCERT not to be republished

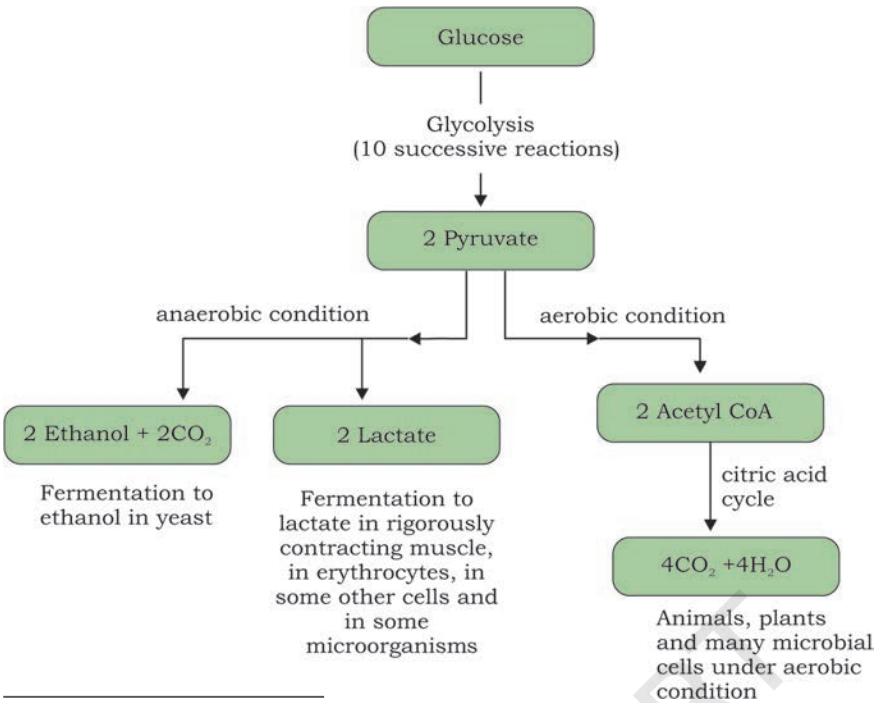


Fig.5.5: Fate of pyruvate

of glycolytic reactions from glucose to pyruvate are similar in most organisms and most types of cells, but the fate of pyruvate is different (Fig. 5.5).

Fermentation (aerobic breakdown of pyruvate)

For glycolysis to continue, NAD^+ which the cells have in limited quantities must be recycled after its reduction to NADH by glyceraldehyde-3-phosphate dehydrogenase. Under anaerobic conditions, the NAD^+ is replenished by the reduction of pyruvate in an extension of the glycolytic pathway.

Homolactic fermentation

Under anaerobic condition, pyruvate is reduced by NADH to lactate by the enzyme lactate dehydrogenase. Lactate dehydrogenase is also an isozyme. The overall process of anaerobic glycolysis in muscle may be represented as,

$$\text{Glucose} + 2\text{ADP} + 2\text{Pi} \longrightarrow 2\text{Lactate} + 2\text{ATP} + 2\text{H}_2\text{O} + 2\text{H}^+$$

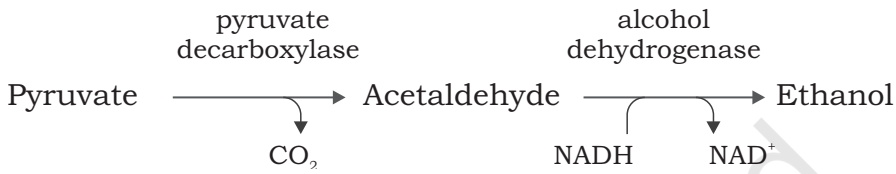
Much of the lactate produced by glycolysis is exported from muscle cell to the liver where it is reconverted to glucose.

Alcoholic fermentation

Under anaerobic conditions in yeast, NAD^+ is regenerated through conversion of pyruvate to ethanol and CO_2.

© NCERT not to be republished

The reaction has two steps. The first reaction is the decarboxylation of pyruvate to form acetaldehyde and CO_2 by the enzyme pyruvate decarboxylase, which contains coenzyme TPP (thiamine pyrophosphate). The second reaction is a reduction of acetaldehyde into ethanol by enzyme alcohol dehydrogenase. In this reaction NADH is oxidised to NAD^+.



Fermentation results in the production of 2 ATP per glucose.

Aerobic breakdown of pyruvate

Under aerobic conditions, pyruvate is transferred from cytosol to mitochondria which is converted to acetyl CoA by oxidative decarboxylation. This is an irreversible reaction, catalysed by a multienzyme complex known as pyruvate dehydrogenase complex (PDH), which is found only in mitochondria.

$$\text{Pyruvate} + \text{CoA} + NAD^+ \longrightarrow \text{Acetyl CoA} + CO_2 + \text{NADH} + H^+$$

The coenzyme and prosthetic groups required in this reaction sequence are TPP, FAD (flavin adenine dinucleotide coenzymes), NAD^+, and lipoamide. This irreversible reaction is the link between glycolysis and citric acid cycle.

5.2.5 Citric acid cycle

The citric acid cycle is also known as **Kreb's cycle** or **Tricarboxylic acid (TCA) cycle** on the name of Hans Kreb who discovered it for the first time (Fig. 5.6). It is a sequence of reactions in mitochondria that oxidise the acetyl moiety of acetyl CoA and reduces NAD^+ that are reoxidised through the electron transport chain, linked to the formation of ATP. The citric acid cycle is the final common pathway for the oxidation of carbohydrates, lipids, and proteins because glucose, fatty acids, and most amino acids are metabolised to acetyl CoA or intermediates of the TCA cycle.

In eukaryotes, the reaction of the citric acid cycle takes place inside mitochondria in contrast with those

© NCERT not to be republished

of glycolysis which occurs in cytosol. The enzymes that catalyse the reactions of TCA cycle are located in the mitochondrial matrix, either free or attached to the inner mitochondrial membrane where the enzymes and coenzymes of electron transport chain are also found.

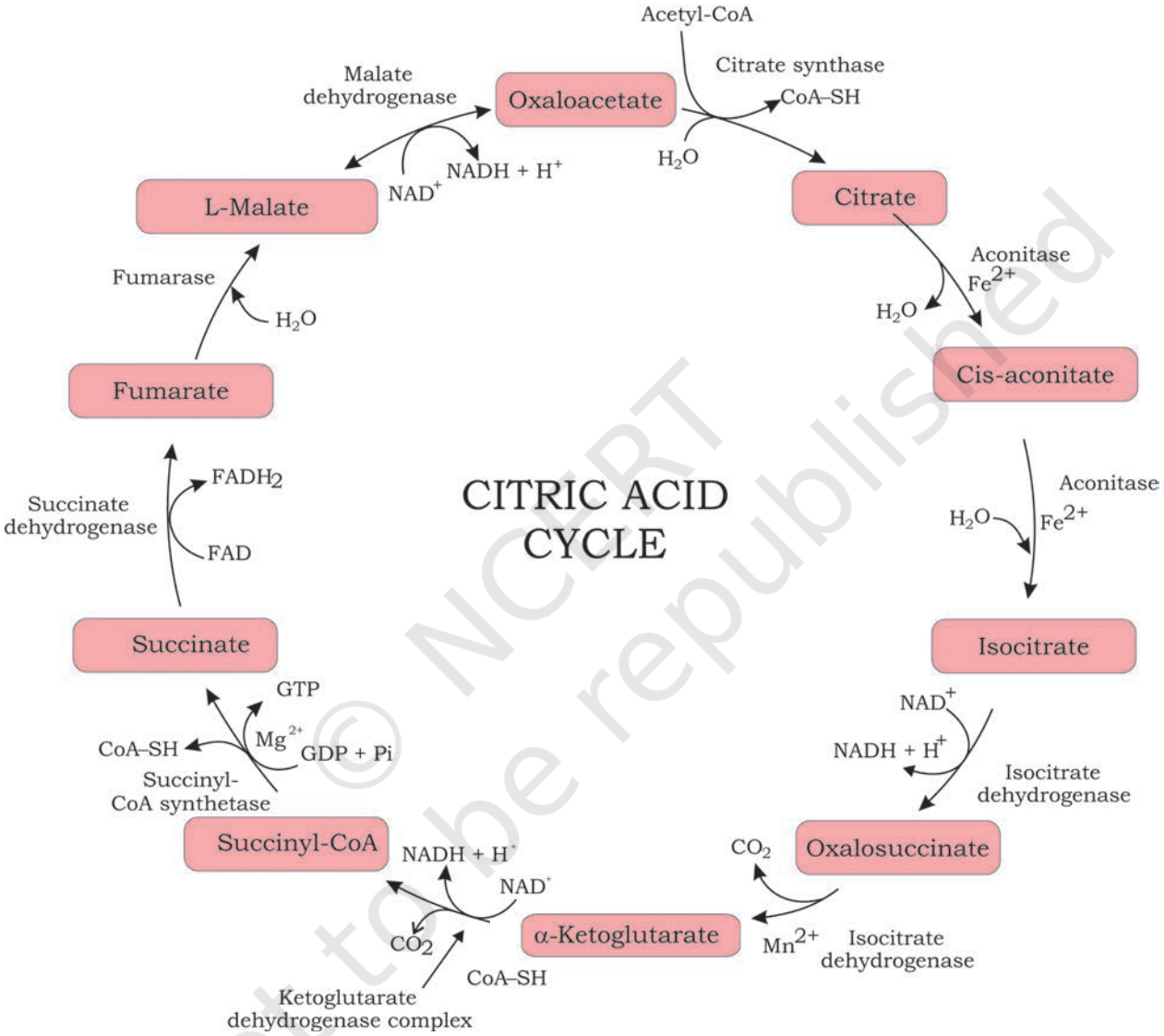


Fig. 5.6: Citric acid cycle

The cycle begins with a condensation reaction between two carbon acetyl CoA and four carbon oxaloacetate to form six carbon citrate by the enzyme citrate synthase (Fig. 5.6). Citrate is isomerised into isocitrate (via the formation of cis-aconitate intermediate) by the enzyme aconitase. Isocitrate undergoes dehydrogenation to form oxalosuccinate by the enzyme isocitrate dehydrogenase.

© NCERT not to be republished

The enzyme isocitrate dehydrogenase is an isozyme present in three forms. One form is found in mitochondria that requires NAD^+ for its activity. The other two forms use $NADP^+$ and are found in both mitochondria and cytosol. In citric acid cycle, oxidation of the isocitrate takes place via NAD^+ dependent enzyme, which results in formation of $NADH+H^+$ along with the product oxalosuccinate. Oxalosuccinate, a transient intermediate remains bound to the enzyme and undergoes decarboxylation to form five carbon α-ketoglutarate. α-ketoglutarate, in a co-enzyme A mediated reaction, undergoes oxidative decarboxylation to form four carbon succinyl CoA by the enzyme α-ketoglutarate dehydrogenase complex. In this step, another NAD^+ get reduced to $NADH+H^+$ and another CO_2 is liberated. This reaction is an irreversible reaction and favors formation of succinyl CoA. Succinyl CoA is converted into succinate by the enzyme succinyl CoA synthetase. This is the only step in a citric acid cycle where, substrate level phosphorylation takes place in which a GDP, is phosphorylated to GTP, which is then converted to ATP. Succinate is then oxidised to form fumarate by the enzyme succinate dehydrogenase. This enzyme contains coenzyme FAD and iron sulfur (Fe-S) protein, the FAD is get reduced to $FADH_2$. In citric acid cycle, succinate dehyfrogenase is the only enzyme which remains attached with the inner membrane of mitochondria, whereas other enzymes remain free in matrix. Fumarate is then converted into malate by enzyme fumarase. Malate is oxidised into oxaloacetate by the enzyme malate dehydrogenase. In this step NAD^+ gets reduced to $NADH+H^+$.

The citric acid cycle harvests the high energy electrons from acetyl CoA and uses these electrons to form NADH and $FADH_2$.

The overall reaction of TCA cycle is

$$\text{Acetyl CoA} + 3NAD^+ + FAD + GDP + Pi + 2H_2O \rightarrow 2CO_2 + 3NADH + 3H^+ + FADH_2 + GTP + CoA$$

5.2.6 Electron transport chain

The reduced coenzymes ($NADH+H^+$ and $FADH_2$) produced during TCA cycle and various other catabolic pathways are oxidised through **electron transport chain (ETC)**

© NCERT not to be republished

located in the inner mitochondrial membrane which are subsequently utilised for production of ATP through ETC. In oxidative phosphorylation the electrons are released during reoxidation of NADH and $FADH_2$ flow through a series of membrane proteins to generate a proton gradient across the membrane. These protons then flow through ATP synthase to synthesise ATP from ADP and inorganic phosphate. Oxygen is utilised as electron acceptor (resulting in the formation of H_2O) at the end of electron transport chain, and for generation of NAD^+, and FAD).

ETC involves chain of four large protein complexes called NADH-Q-oxidoreductase (complex I), succinate-Q-reductase (complex II), Q-cytochrome C oxidoreductase (complex III), and cytochrome C oxidase (complex IV) (Fig. 5.7). These electron carriers are large transmembrane protein complexes with multiple oxidation-reduction centers such as quinones, flavins, iron sulfur clusters, heme and copper ions. Succinate-Q-reductase (complex II) in contrast with the other three complexes does not pump protons. Electrons are carried from complex I to complex III by the reduced form of coenzyme Q, also known as ubiquinone. Ubiquinone is quinone, which is lipid (hence hydrophobic) and diffuses quickly within the inner mitochondrial membrane. The electrons from $FADH_2$ are generated through succinate-Q-reductase (complex II), which are then transferred to cytochrome-C-oxidoreductase (complex III). Then electron from complex III is transferred to cytochrome-C-oxidase (complex IV), which is the final component in the ETC. This component catalyses the reduction of O_2 into H_2O.

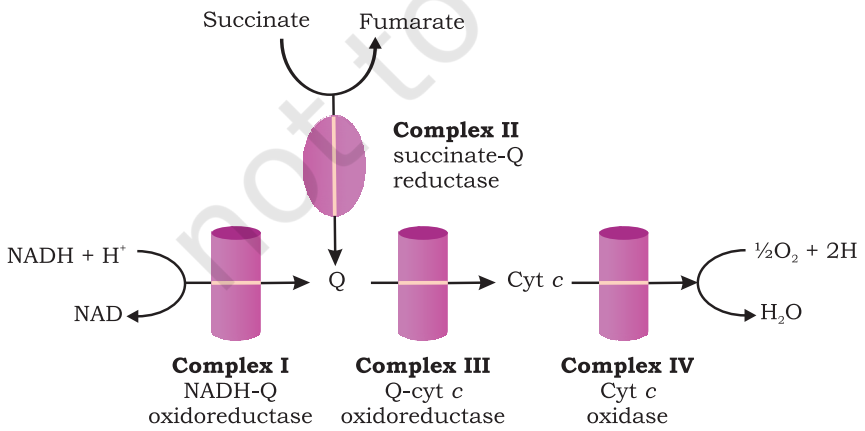


Fig.5.7: The four complexes of electron transport chain

© NCERT not to be republished

Oxidative phosphorylation (ATP synthesis)

The flow of electron from NADH or $FADH_2$ to O_2 through four protein complexes located in the mitochondrial inner membrane, lead to the pumping of protons out of the mitochondrial matrix side to the cytosolic side of the inner mitochondrial membrane (Fig. 5.8). The resulting uneven distribution of protons generates a proton motive force (pmf), which consists of a pH gradient (matrix side basic) and a trans membrane electrical potential (matrix side negative). Synthesis of ATP takes place when protons flow back to the matrix side through an enzyme complex called ATP synthase.

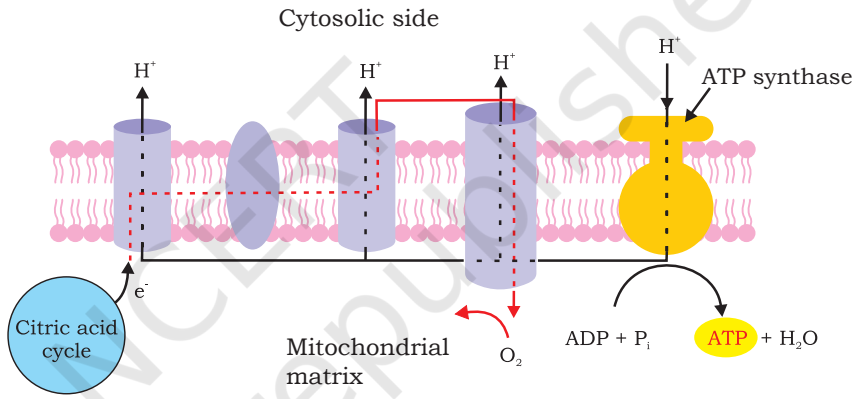


Fig.5.8: Proton pumping and pH gradient

The flow of electrons through NADH-Q-oxidoreductase, Q-cytochrome-c-oxidoreductase, and cytochrome-c-oxidase develop a gradient enough to synthesise 1, 0.5 and 1 molecule of ATP, respectively. Hence, 2.5 molecules of ATP are synthesised per molecule of NADH oxidised in the mitochondrial matrix, whereas only 1.5 molecule of ATP are made per molecule of $FADH_2$ oxidised because its electrons enter the electron transport chain at QH_2 (complex III).

The citric acid cycle along with oxidative phosphorylation takes place in mitochondria and is the major source of ATP in aerobic organisms.

5.2.7 Photosynthesis

All the energy utilised by biological system come from solar energy, which is trapped and converted to chemical energy in the process of photosynthesis. The basic reaction

© NCERT not to be republished

of photosynthesis can be represented as:

$$6CO_2 + 12H_2O \xrightarrow[\text{Chlorophyll}]{\text{light}} C_6H_{12}O_6 + 6O_2 + 6H_2O \text{ .....(1)}$$

where, $C_6H_{12}O_6$ represents carbohydrate. Photosynthesis in plants and some photosynthetic bacteria take place in chloroplasts. The energy of light captured by various photoreceptor pigment molecules, which includes chlorophyll a, xanthophylls, carotenoids and other chlorophyll, found in chloroplasts, is used to generate high energy electrons with great reducing potential. These electrons are utilised in the production of ATP and $NADPH + H^+$ through a series of reactions called the **light reactions** of photosynthesis. ATP and $NADPH+H^+$ formed in the process of light reaction reduce CO_2 and convert it into 3-phosphoglycerate by a series of reactions called **Calvin cycle** or **dark reactions**.

5.2.7.1 Light Reactions

Light reactions are also called 'Photochemical' phase of photosynthesis. This phase includes the following:

1. Capturing the solar energy,
2. Splitting of water and release of oxygen, and
3. Formation of higher-energy organic intermediates, i.e. ATP and $NADPH + H^+$.

The light reactions occur in the thylakoid membranes of chloroplast. The thylakoid membrane comprises the major energy transducing machinery such as light harvesting proteins, reaction centers, electron transport chains, and ATP synthase.

In membrane system of chloroplast, there are two types of light-harvesting protein complexes, namely, **photosystem I (PSI)** and **photosystem II (PSII)** (Fig. 5.9) whose composition is quite variable. Typically, PSI contains 13 polypeptide chains, about 60 chlorophyll molecules, a quinone (vitamin K) and three 4Fe-4S clusters. Total molecular mass of PSI is 800 kDa. PSII is slightly less complex with about 10 polypeptide chains, 30 chlorophyll molecules, a nonheme iron ion and four manganese ions.

The two photosystems, PSI and PSII, respond to light at slightly different wavelengths. PSI responds to the wavelength of 700 nm, whereas chlorophyll a at PSII responds to the wavelength of 680 nm.

© NCERT not to be republished

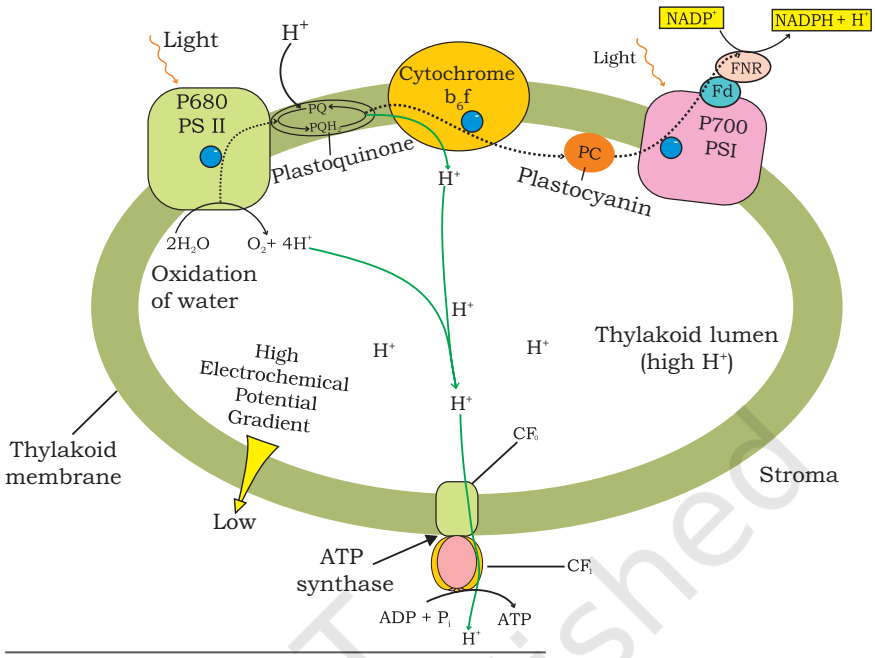


Fig. 5.9: Light reactions of photosynthesis

Under normal conditions, PSII is activated first. Capture of red light at 680 nm by PSII brings an electron from its ground energy state to an excited state. The excited electron enters the **Electron Transport Chain (ETC)** by acceptor molecule pheophytin (chlorophyll with two H^+ ions in place of Mg^{2+} ion), which transfers electron to plastoquinone (PQ), from PQ to cytochrome (Cyt b_6f) complex, from cytochrome b_6f complex to plastocyanin (PC), from PC to PSI. PSI accepts electron from the PC only when the electron in its reaction centre gets excited by the absorption of light at 700 nm. Acceptor molecule (a modified chlorophyll) of $P700^+$ transfer its electron to a phylloquinone and from phylloquinone to soluble ferredoxin (Fd) through a series of membrane bound iron-sulphur protein, which finally reduces $NADP^+$ into NADPH + H^+ in the stroma, in the presence of the enzyme ferredoxin NADP-reductase (FNR).

Transfer of electron from PSII (P680) to pheophytin results in positive charge on it ($P680^+$), which is a very strong oxidant and receives an electron from water present in the lumen of thylakoid. Two molecules of water are oxidised to generate one molecule of O_2 for every four electrons sent from PSII to PSI through ETC. The splitting

© NCERT not to be republished

of water is catalysed by a protein complex, the oxygen evolving complex, located on the luminal surface of the thylakoid membrane.

$$2H_2O \xrightarrow{O_2 \text{ evolving complex}} 4H^+ + 4e^- + O_2 \quad \ldots\ldots(2)$$

We can look at the reaction (2). It reveals that water oxidises and provide hydrogen ion as proton. If we carefully observe at the Fig. 5.9, we can find that during transfer of electron to plastoquinone, it accepts proton from stroma and transfer it to the lumen of thylakoid through cytochrome b_6f. This input of proton into the lumen of thylakoid results in higher concentration of protons inside. This results into development of proton gradient across the membrane i.e., very high electrochemical potential gradient in lumen and very low in stroma side of the thylakoid membrane. As a result of this proton motive force (pmf), protons tend to re-enter into the stroma through proton channel (CF_0) of the ATP synthase. ATP is synthesised on the CF_1 subunit of the ATP synthase by using ADP and inorganic phosphate (Pi). In this way both NADPH + H^+ and ATP are synthesised by non-cyclic electron flow, which we call **non-cyclic photophosphorylation**.

In certain cases, where only PSI works, it follow the cyclic flow of electron. The PSI is present in the

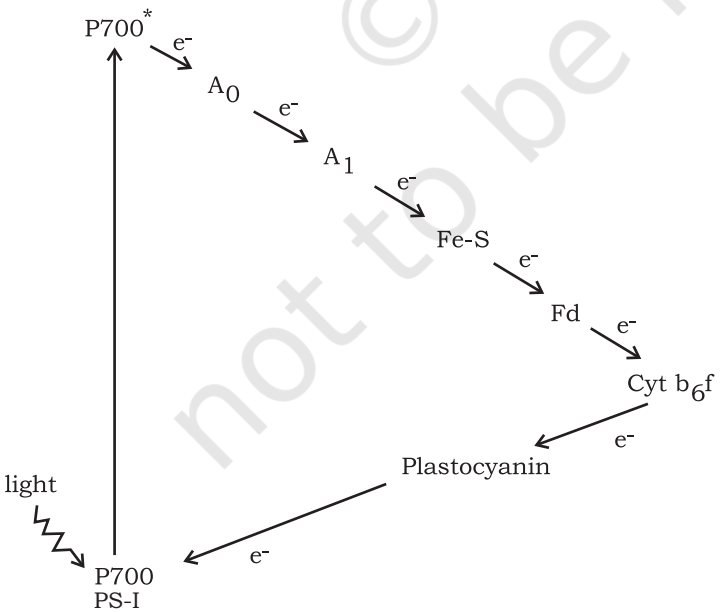


Fig. 5.10: Cyclic photophosphorylation

© NCERT not to be republished

stromal lamellae or on the edge of the grana of thylakoid membrane. When the reaction center of PSI absorbs light at 700 nm, its photoexcited electrons are accepted by a modified chlorophyll molecule (A_O). A_O transfers its electron to phylloquinone (A_1) which diffuses into thylakoid membrane and binds to cytochrome b_6f complex (Fig. 5.10). Phylloquinone is coupled to proton pumping into thylakoid membrane from stroma side to create **electrochemical proton gradient**. ATP synthesis occur when proton back-flow down their electrochemical gradient through ATP synthase. Cytocrome b_6f transfers its electron to PSI through plastocyanin. Thus, photoexcited electron from PSI again come in their original state through electron transport chain; so it's called **cyclic electron flow**. In cyclic electron flow only ATP is produced, not NADPH + H^+ and O_2. The process of formation of ATP due to light induced cyclic flow of electron is called **cyclic photophosphorylation**.

5.2.7.2 Dark reactions or Calvin Cycle

The dark reactions of photosynthesis take place within the stroma of chloroplast in plants and cytosol in cyano-

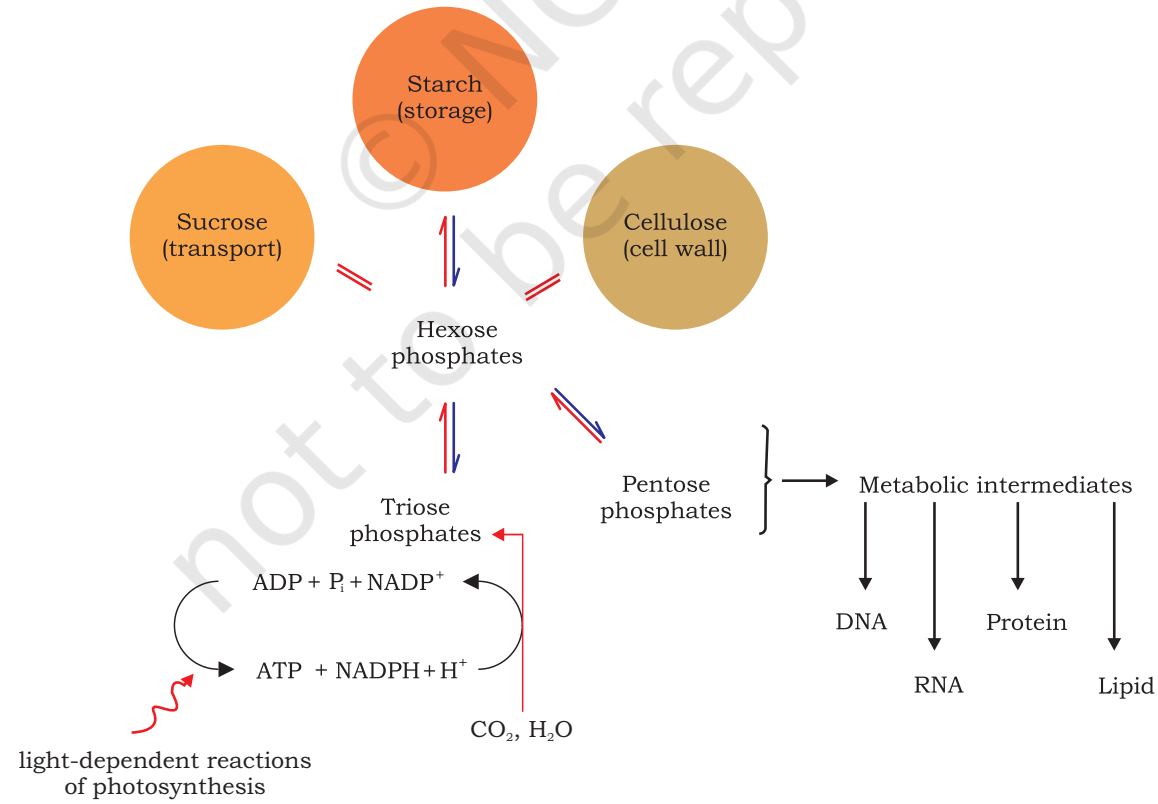


Fig. 5.11: Carbon assimilation

© NCERT not to be republished

bacteria. The stroma contains the soluble enzymes that utilise the NADPH+H^+ and ATP synthesised by the thylakoids to convert CO_2 into sugar. The energy and reducing power generated in the form of ATP and NADPH+H^+, respectively, by light-dependent reactions of photosynthesis are utilised by plants to synthesise carbohydrate. In this process, CO_2, which is the sole source of carbon assimilate into trioses and hexoses which lead to the synthesis of sugar. The process is called **CO_2 assimilation**, or **CO_2 fixation** (Fig. 5.11).

The CO_2 fixation operates in a cyclic pathway, known as **Calvin cycle** as it was elucidated in 1950 by **Melvin Calvin** and his co-workers. The Calvin cycle takes place in the stroma of chloroplasts. We can understand CO_2 assimilation in three stages, stage I, stage II and stage III (Fig. 5.12).

Stage I: Fixation of CO_2 into 3-phosphoglycerate

In this stage, gaseous carbon dioxide is fixed into stable organic intermediate. CO_2 get condensed with five carbon acceptor ribulose1, 5-bisphosphate to form two molecules of 3-phosphoglycerate (PGA); this carboxylation reaction is catalysed by enzyme ribulose1, 5-bisphosphate carboxylase/oxygenase (also known as **RuBisCO**). The enzyme **RuBisCO** possesses both carboxylase and oxygenase activities. Three molecules of CO_2 are fixed to three molecules of ribulose1, 5-bisphosphate to form six molecules of 3-phosphoglycerate (PGA) (Fig. 5.12). Since first product of this stage of dark reaction in 3-phosphoglycerate, which is a three carbon molecule, therefore, this cycle is also known as C-3 cycle.

Stage II: Reduction of 3-phosphoglycerate to glyceraldehyde 3-phosphate

The six molecules of 3-phosphoglycerate are reduced to six molecules of glyceraldehyde-3-phosphate. During this process six molecules of ATP are utilised to convert six molecules of 3-phosphoglycerate into six molecules of 1,3-bisphosphoglycerate. Subsequently six molecules of NADPH+H^+ are consumed to reduce six molecule of 1,3-bisphosphoglycerate into six molecules of glyceraldehyde-3-phosphate.

© NCERT not to be republished

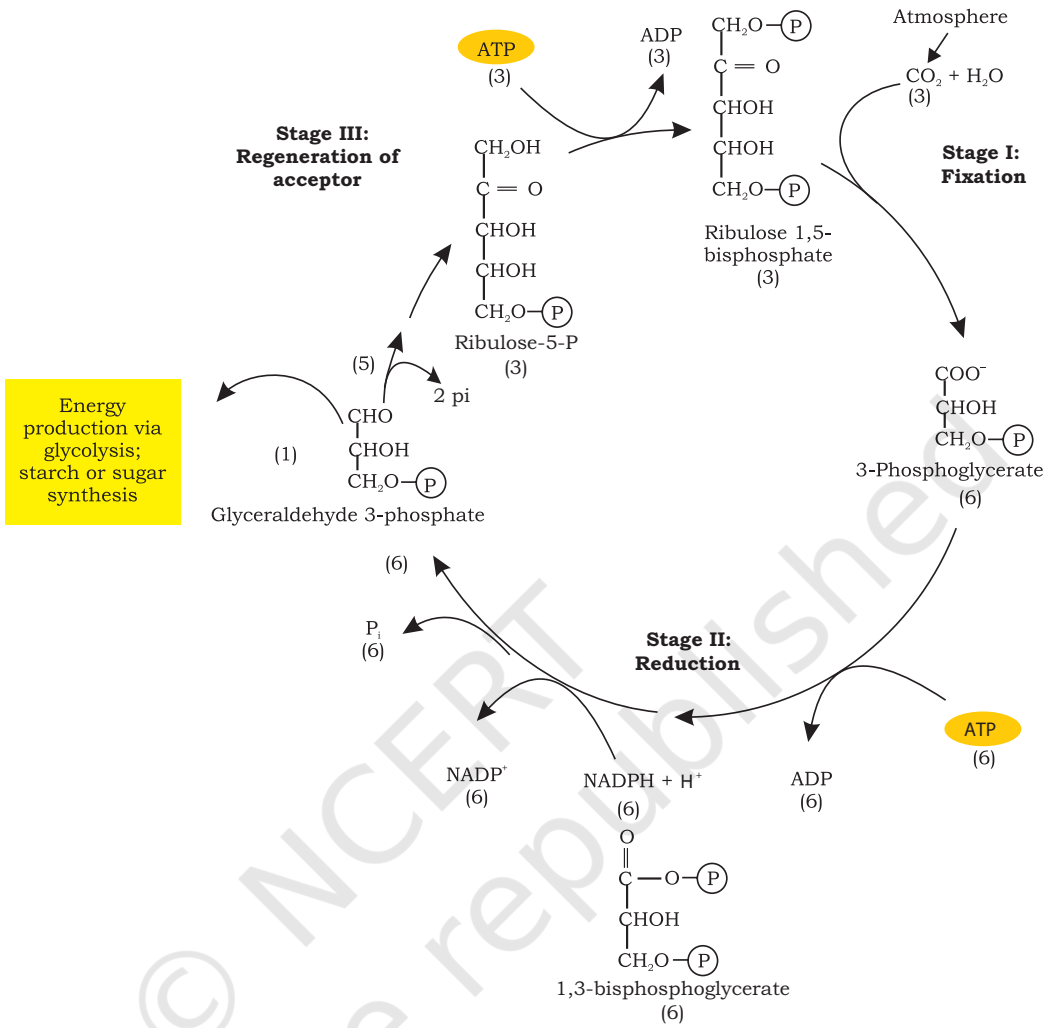


Fig. 5.12: The three stages of Calvin cycle

Stage III: Regeneration of ribulose1, 5-bisphosphate from triose phosphates

In this stage, out of six molecules of glyceraldehyde-3-phosphate, one molecule is diverted for energy production via glycolysis, starch or sugar synthesis, while remaining five molecules of glyceraldehyde-3 phosphate (15C) are converted to three molecules of ribulose-5-phosphate (15C) with concomitant release of two inorganic phosphate (2Pi). These three molecules of ribulose-5-phosphate (15C) are converted to three molecules of ribulose-1,5-bisphosphate. During this conversion three molecules of ATP are utilised releasing three molecules of ADP.

© NCERT not to be republished

Uptill now, the photosynthesis pathway described above is followed by a majority of plants, which are called C3 plants, and the pathway followed by them is known as C3 pathway. However, some plants follow a different pathway, which is called the C4 pathway, described in the following section.

5.2.7.3 C_4 pathway

The C4 pathway of photosynthesis is followed by certain plants such as maize, sorghum and sugarcane, which have approximately 50% higher photosynthesis efficiency than those of C3 plants. C4 plants have a distinct leaf anatomy from that of C3 plants, which is known as Kranz anatomy. In Kranz anatomy the mesophyll cells are clustered around bundle sheath cells in a ring-like fashion and the number of chloroplast in bundle sheath cells is more than those in mesophyll cells.

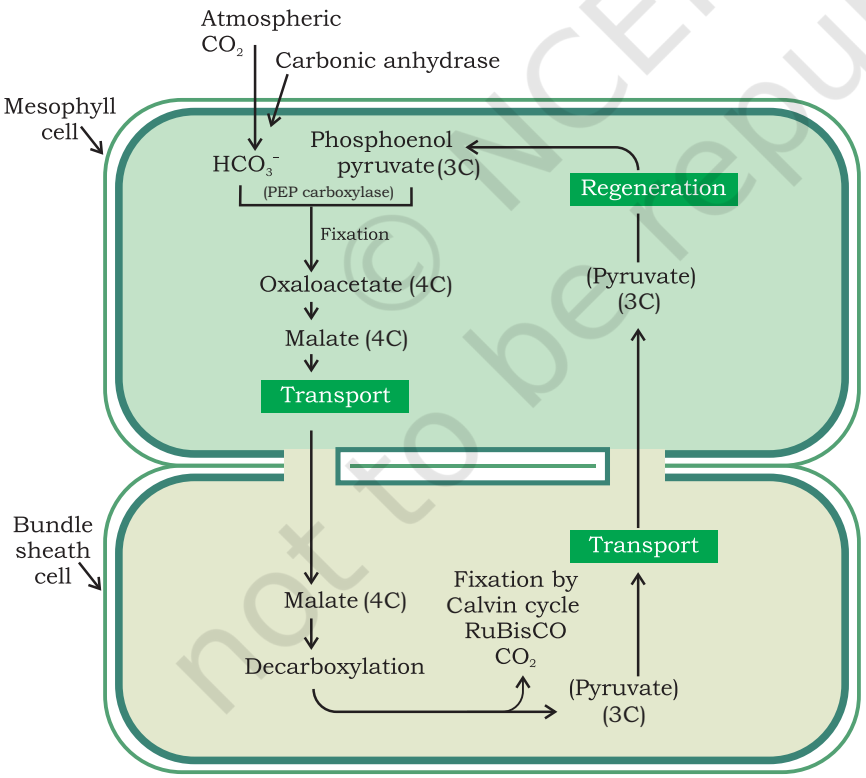


Fig. 5.13: C_4 cycle

© NCERT not to be republished

Now let us study the pathway as given in Fig. 5.13. Mesophyll cells in leaves of C_4 plants has 3 carbon molecule, phosphoenol pyruvate (PEP) which is acceptor of CO_2. PEP carboxylase enzyme is responsible for the fixation of CO_2 and formation of C_4 acid that is oxaloacetate (OAA) in the mesophyll cell. Oxaloacetate forms malate by NADP-malate dehydrogenase enzyme and transports it to bundle sheath cells. In bundle sheath cells malate is broken down to release CO_2 and a 3-carbon molecule pyruvate. Pyruvate is transported to mesophyll cells which regenerate to form phosphoenol pyruvate. Since oxaloacetic acid, which is a four carbon molecule, is the product of first stage of this reaction, therefore, this cycle is called C_4 cycle.

Carbon dioxide released due to decarboxylation of malate is accepted by RuBisCO in bundle sheath cell for C_3 pathway or Calvin cycle. Bundle sheath cells are rich in an enzyme RuBisCO, but lack PEP carboxylase. Thus C_3 pathway is a basic pathway and is common to all plants. In mesophyll cells of the C_3 plants PEP carboxylase is absent. In C_3 plants, RuBisCO is present in mesophyll cells of the leaves which is first CO_2 acceptor, while in C_4 plants RuBisCO is present in bundle sheath cells and first CO_2 acceptor is PEP carboxylase, which is found in mesophyll cells.

5.3 CELL CYCLE

All living organisms start their life from a single cell. But where do the new cells come from? One of the most important characteristics of a cell is its ability to **grow** and **divide** (or reproduce) in suitable environment. You might ask, why cells need to divide? Such a question can be answered in terms of what happens if they do not? The answer is—they die. **Cell division** is common to all living organisms and is essentially a mechanism by which cells grow and divide, giving rise to a new cell population. As a cell divides, it gives rise to two daughter cells which possess the same genetic makeup as the parent cell. These daughter cells further give rise to new cells by undergoing reproduction. You must now appreciate the formidable ability of cells to divide by considering the fact that we have trillions of cells in our body despite starting our lives as a single cell.

5.3.1 Phases of cell cycle

There are a series of tasks which a cell must accomplish

© NCERT not to be republished

in order to divide. These include: growth, replication of genetic material and physical splitting of the cell into two daughter cells. All these events are themselves under strict genetic control. A typical eukaryotic cell takes about **24 hours** to divide. However, this duration is not absolute, and can vary from organism to organism.

A cell cycle is divided into two major phases: interphase and mitotic phase.

1. Interphase—This phase is basically characterised by growth of the cell, followed by replication of its DNA content. It represents the phase between two successive mitotic phases (Fig. 5.14).

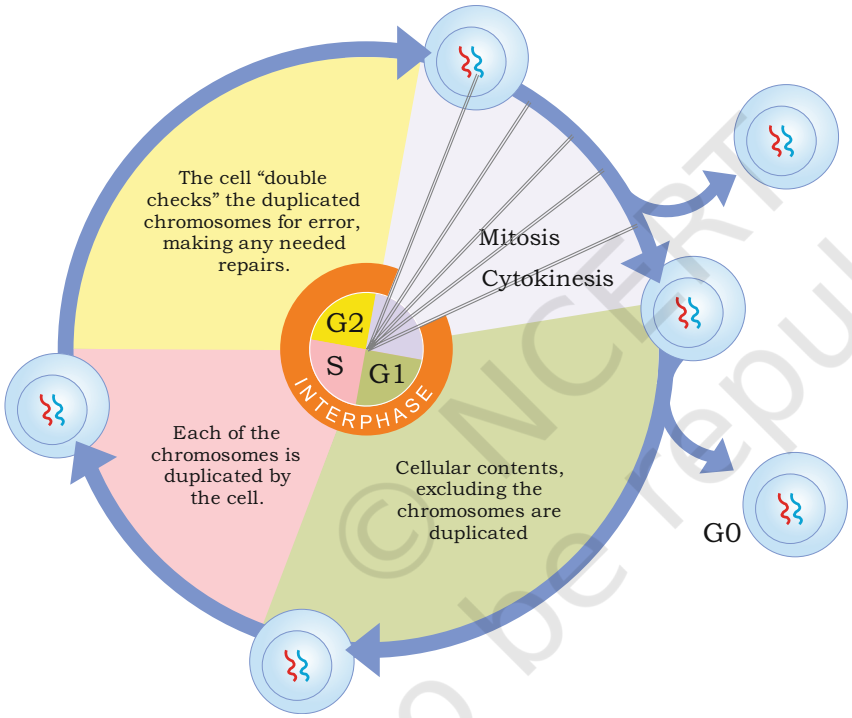


Fig. 5.14: A schematic diagram of cell cycle

2. Mitotic (M) Phase—During M-phase, the cell begins to separate its DNA content (which it had replicated during interphase) as well as cytoplasm, making sure that both the daughter cells receive exactly the same amount of DNA at the end of cell cycle.

A detailed discussion of both the abovementioned phases is given in the following sections.

Interphase

A cell which has just entered into the cell cycle must grow and carry out DNA replication in an orderly manner.

© NCERT not to be republished

This is accomplished during the interphase of cell cycle. Interphase is also sometimes referred to as the resting phase because it is the time when the cell is preparing itself to undergo replication. The interphase is further divided into three phases, namely **G1 (Gap 1) phase, S (Synthesis) phase,** and **G2 (Gap 2) phase** (Fig. 5.14).

1. ***G1 Phase***— During the first gap phase, the cell is metabolically active and grows continuously larger but does not replicate its genetic content. G1 phase is essentially the first gap which occurs between mitosis and synthesis of DNA. The growth of the cell occurs largely due to the accumulation of RNA, organelles, and other molecular building blocks.

2. ***S Phase***— This phase is characterised by the synthesis of DNA in the nucleus. During S phase, the amount of DNA in the cell doubles. Therefore, if initially the DNA content in the cell was 2C (the amount of DNA of genome), at the end of S phase the amount increases to 4C. It is important to note that the chromosome number in the cell remains unchanged at the end of this phase, i.e., a diploid (2n) cell will remain diploid. Microtubule-organising structure known as the **centriole** is also duplicated in the cytoplasm during S-phase. Centrioles are required for the proper segregation of chromosomes.

3. ***G2 Phase***— This phase is marked by the synthesis of more proteins and organelles which further enhances cell growth. The cell, at this stage, is prepared to enter the M phase.

In the body of an organism, not all cells are equally capable of undergoing division. In certain cell types, for example somatic cells, mitotic division occurs on a regular basis. However, in other cell types, the cells are incapable of undergoing division. These cells withdraw themselves from the cell cycle by exiting the G1 phase and enter an inactive stage called **quiescent stage (G0)**. Such cells are also referred to as **differentiated cells**, for example - cardiac cells, neurons, etc.

Mitotic (M) Phase

Mitosis is composed of two broad phases: karyokinesis and cytokinesis. During M phase, the cell condenses its chromosomes and segregates them equally into two daughter cells. Since the number of chromosomes in the

© NCERT not to be republished

parent and daughter cells is the same, it is also referred to as **equational division**. Karyokinesis is further divided into the following four stages (Fig. 5.15):

1. Prophase—Mitosis begins with the condensation of chromatin in the prophase (Fig. 5.15). The chromatin material which is lying diffused in the nucleus during interphase can be seen as long threads. However, during the prophase of mitosis, these chromatin threads begin to condense. The condensed form of chromatin is known as chromosome. This phase is also marked by the movement of centrioles, which were duplicated during the S phase, towards the opposite poles of the cell. The movement of centrioles towards the opposite poles initiates the assembly of mitotic spindle. Importantly, during the prophase, a number of cell organelles and structures begin to disappear. These include the Golgi apparatus, endoplasmic reticulum, nucleolus and nuclear membrane.

2. Metaphase—By this stage, the disintegration of nuclear envelope and condensation of chromosomes is completed. The chromosomes, at this stage, can be easily viewed under the microscope and their morphology can be studied (Fig. 5.15). Two sister chromatids, held together at the **centromere** is a general feature of a metaphase chromosomes. Centromeres consist of small disc-shaped structure at their surface known as **kinetochores**, which serve as the sites of attachments of chromosomes to the spindle fibers. A general feature of metaphase is that all the chromosomes come to lie at the equator with one chromatid of each chromosome connected by its kinetochore to spindle fibers from one pole, and its sister chromatid connected by its kinetochore to spindle fibers from the opposite pole. The plane of alignment of the chromosomes during metaphase is referred to as the **metaphase** plate.

3. Anaphase—During this stage, chromosomes arranged at the metaphase plate are eventually pulled apart towards the opposite poles.

4. Telophase—This is the final stage of M phase, marked by the reformation of the nuclear envelope. By this stage, all the chromosomes which were split apart during the anaphase, have reached their respective poles. These chromosomes then begin to decondense.

© NCERT not to be republished

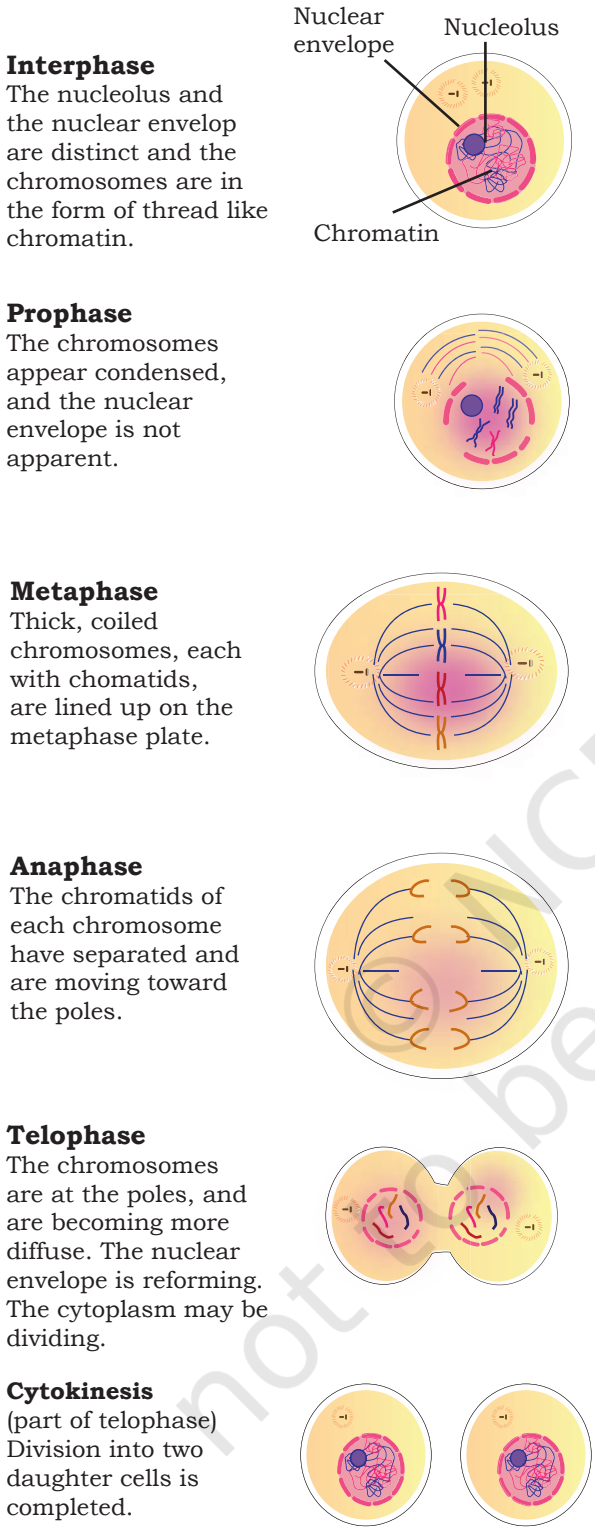


Fig. 5.15: The various stages of M phase

5.3.2 Cytokinesis

The division of cytoplasm of a parent cell occurs as a result of a physical process called **cytokinesis** (Fig. 5.15). This process results in the separation of a parent cell into two daughter cells. Even before cytokinesis occurs, division of the nucleus (karyokinesis) occurs. Cytokinesis begins with the formation of a furrow in the plasma membrane. As the furrow deepens, it ultimately joins at the centre of the cell forming a contractile ring, dividing the cytoplasm equally in two. The contractile ring is made up of actin filaments. However, in plant cells, the situation is different because plant cells are surrounded by an inextensible cell wall. Therefore, to successfully carry out cell division, plant cells synthesise new cell wall during cytokinesis. The formation of the new cell wall begins with formation of a precursor, called cell plate. Cytokinesis is characterised by the distribution of cell organelles equally into two daughter cells.

5.3.3 Meiosis

Meiosis refers to the production of haploid gametes from diploid cells. These gamete cells ultimately fuse during the fertilisation phase of sexual reproduction to produce diploid offspring. Meiosis is also seen during gametogenesis in plants. During meiosis, the number of chromosome is reduced to half. There is only a single interphase in meiosis, which is followed by two nuclear divisions. The two nuclear division cycles are known as **Meiosis I** and **Meiosis II** (Fig. 5.17).

The first meiotic division (Meiosis I) is also known as **reductional division** in which the number of chromosomes is reduced to half, while meiosis II is

© NCERT not to be republished

similar to mitosis and referred to as **equational division** in which the number of chromosomes remains the same. At the end of a meiotic cycle four haploid cells are produced. Meiosis differs from mitosis in two crucial respects:

(i) Number of chromosomes in mitosis remains the same as the parent cell, while in meiosis the number of chromosomes is halved.

(ii) Recombination during meiosis results in shuffling of genes between chromosomes in each pair.

Meiosis I and II are divided into the following stages: Prophase, Metaphase, Anaphase and Telophase. However, in order to distinguish the two meiotic cycles, the subphases of meiosis I and meiosis II are referred to as Prophase I, Metaphase I, Anaphase I and Telophase I for meiosis I, and Prophase II, Metaphase II, Anaphase II and Telophase II for meiosis II.

Prophase I: Prophase I is the longest phase of meiotic cycle. During prophase I, a number of activities occurring between the chromosomes are required for their proper segregation in the subsequent stages. Prophase I is further divided into the following substages (Fig. 5.16):

(i) Leptotene—During this stage, the chromosomes are in the form of thin threads visible under the light microscope. Each chromosome consists of two sister chromatids.

(ii) Zygotene— The pairing of homologous chromosomes takes place during this stage. The synaptonemal complex is assembled which facilitates the pairing of homologous chromosomes. The complex formed by a pair of synapsed homologous chromosomes is called a bivalent or tetrad.

(iii) Pachytene— This stage is characterised by the appearance of recombinational nodules. Recombinational nodules are the sites on chromosomes where **crossing over** takes place between non-sister chromatids of homologous chromosomes. Crossing over is the exchange of genetic material between homologous chromosomes. It results in recombination of genetic material and is a source of variation among organisms of same species.

© NCERT not to be republished

(iv) ***Diplotene***— This stage is marked by the degeneration of synaptonemal complex and separation of homologous chromosomes except at the crossover sites known as **chiasmata** (Singular: Chiasma). In oocytes of some vertebrates, diplotene can last for months or years.

(v) ***Diakinesis***— During diakinesis, the chiasmata are terminalised. The chromosomes become fully condensed and the meiotic spindle is assembled to mediate the separation of homologous chromosomes. By the end of this stage, the nuclear envelope breaks down and the nucleolus disappears.

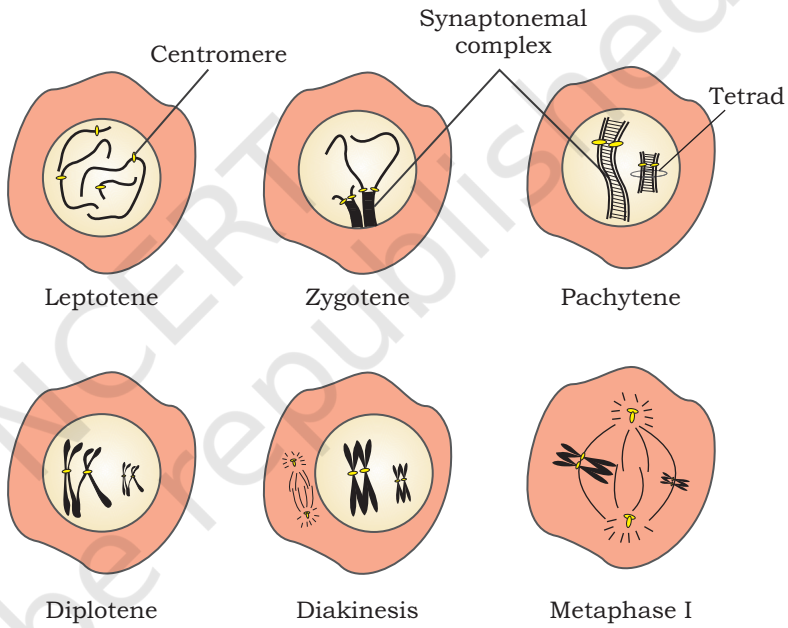


Fig. 5.16: The various stages of Prophase of meiosis I

Metaphase I

This stage is marked by the alignment of bivalents on the equatorial plate. This alignment is facilitated by the attachment of microtubules from the opposite poles of the spindle to the kinetochore of homologous chromosomes (Fig. 5.17).

Anaphase I

Separation of homologous chromosomes occur during anaphase of meiosis I. However, the sister chromatids remain attached to each other at their centromeres.

Telophase I and cytokinesis

During this stage, the nuclear envelope and nucleolus

© NCERT not to be republished

reappear which is closely followed by cytokinesis. Generally, there is a gap between two meiotic divisions called interkinesis. This phase is short-lived and does not involve another round of DNA replication.

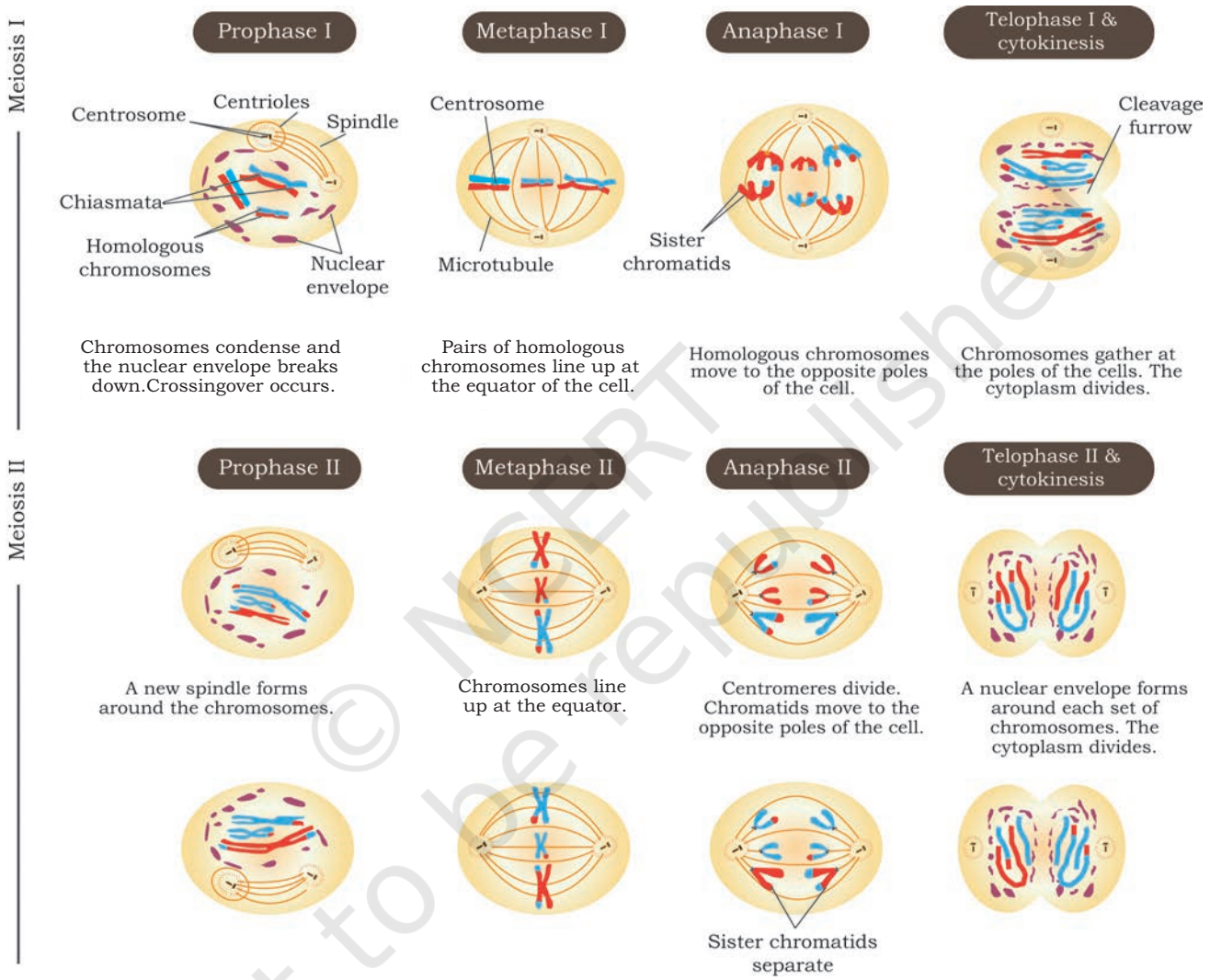


Fig. 5.17: The different stages of Meiosis I and Meiosis II

Meiosis II

Prophase II— During prophase II, there is breakdown of nuclear envelope and disappearance of nucleolus. The compaction of chromosomes occurs again.

Metaphase II— Attachment of chromosomes occurs at the centromeres containing kinetochores. The chromosomes are arranged at the equator.

© NCERT not to be republished

Anaphase II— The sister chromatids of each chromosome separate out and move towards opposite poles.

Telophase II and cytokinesis— This phase marks the end of meiotic cycle. The two sets of chromosomes get enclosed by a nuclear membrane as it reforms, followed by cytokinesis. The final result of meiosis is the formation of four haploid cells.

5.3.4 Significance of meiosis

(i) It ensures the same chromosome number (n) in all the sexually reproducing organisms.

(ii) It helps to restrict the number of chromosomes and maintains stability of the species.

(iii) Crossing over which occurs between the homologous chromosomes during meiosis is a significant source of genetic variations among the offspring.

(iv) All four sister chromatids of homologous chromosomes segregate and go to four different daughter cells. This makes the four daughter cells genetically different.

Table 5.1: How is Meiosis Different from Mitosis?

S.No.	Mitosis	Meiosis
1.	Mitosis occurs in both sexually as well as asexually reproducing organisms.	Meiosis occurs only in sexually reproducing organisms.
2.	Mitosis takes place in the somatic cells of the body.	Meiosis takes place in the germ cells.
3.	During mitosis, the cell undergoes only one nuclear division.	During meiosis, the cell undergoes two nuclear divisions.
4.	DNA replication takes place at interphase.	DNA replication takes place at interphase I but not at interphase II.
5.	Prophase is comparatively simple.	Prophase I is divided into further sub-phases.
6.	Synapsis does not occur in mitosis.	Synapsis of homologous chromosomes occur at prophase.
7.	Crossing over between sister chromatids does not occur during mitosis.	Crossing over occurs between non-sister chromatids of homologous chromosomes.
8.	In daughter cells number of chromosomes is equal to the mother cell.	The daughter cells get half chromosomes to that of the mother cell.
9.	Mitosis results in formation of two daughter cells	Meiosis results in formation of four daughter cells.

© NCERT not to be republished

5.4 Programmed Cell Death (Apoptosis)

During the embryonic stage of development, our fingers are joined together, giving our hands a web-like appearance. As development continues, gaps begin to appear between our fingers. If it were not so, we would have webbed-hands and would not be able to hold things firmly. But how are these gaps created? The answer is that the cells between our fingers undergo apoptosis.

Apoptosis (also known as **Programmed Cell Death**) is a property of the cells which enables them to die during development. It is a highly controlled, energy-dependent process and does not occur randomly. It means that once a cell is committed to undergo apoptosis, it cannot reverse its fate. Not only is apoptosis useful in normal development, but also plays a major role in protecting us from diseases including viral infections and cancers. In certain cases where apoptosis does not work efficiently, unregulated growth and division of cells may occur which could result in the formation of cancer.

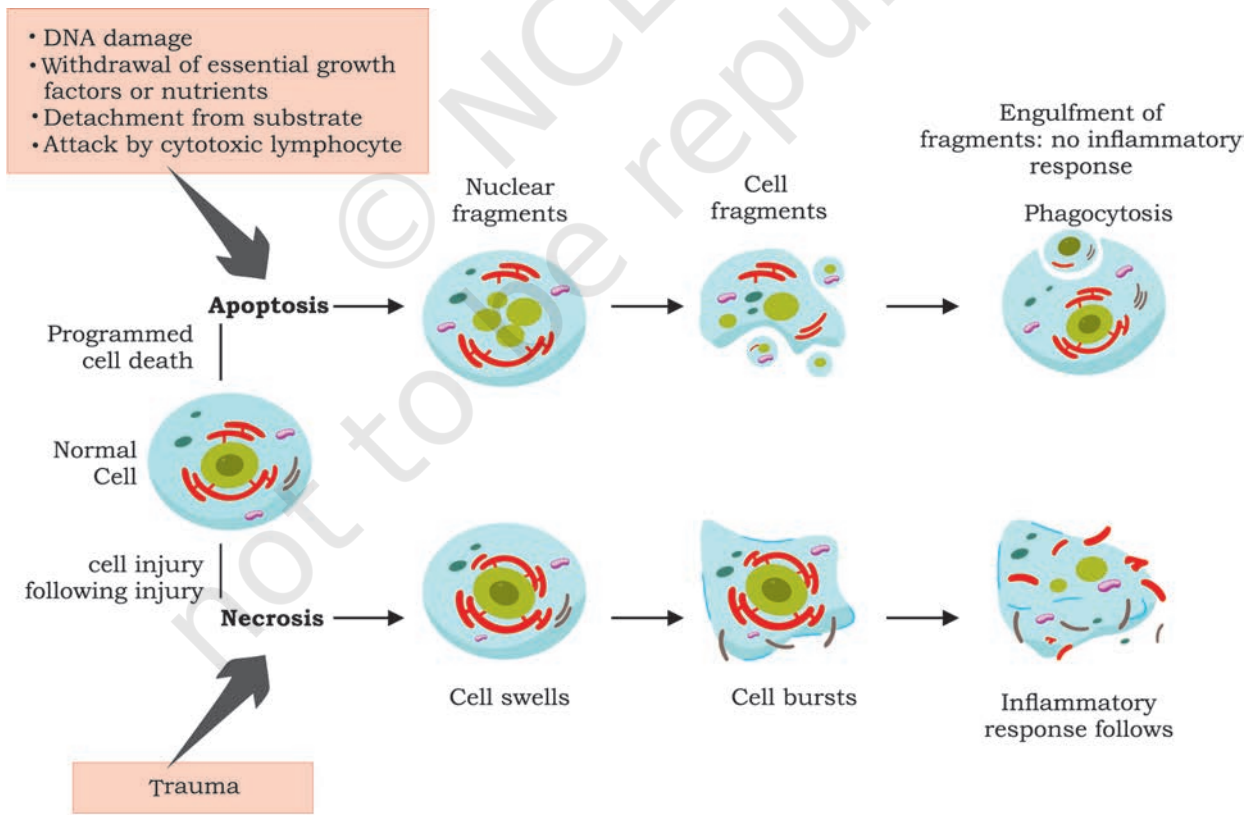


Fig. 5.18: The mechanism of Apoptosis and Necrosis

© NCERT not to be republished

Apoptosis is characterised by certain morphological and physiological changes in the cell. These include DNA fragmentation, blebbing of the plasma membrane, breakdown of nuclear envelope and increase susceptibility of DNA to deoxyribonuclease (DNAase). Eventually, the entire cell splits up into small membrane-enclosed fragments which are ultimately phagocytosed by the nearby macrophages (Fig. 5.18). It is important to note that apoptosis is not the only mechanism by which cells die. Sometimes, when there is a physical injury, an immune response is triggered which results in inflammation of the injured area. In such conditions, the cells undergo **necrosis**, which results in unregulated digestion of cell fragments. Essentially, necrosis is a traumatic, unplanned death of cells which occurs as a result of injuries or exposure to toxic substances. Unlike apoptosis, necrosis is not a regulated process and does not involve DNA fragmentation or blebbing of plasma membrane. The content of a cell which undergoes necrosis is released into the surrounding and causes inflammation.

5.5 CELL DIFFERENTIATION

You are aware that a multicellular organism consists of trillions of cells having exactly the same DNA. Yet, it is observed that a group of cells is different from other group of cells in terms of their structural and functional roles. For instance, a neuron is specialised to conduct impulse whereas RBCs perform the function of oxygen transport. How are these distinctions created? The answer lies in the way genes are expressed in different cells. In other words, the unique combination of genes which are differentially expressed at a given time in different cells dictates their structure and function.

As development of an embryo proceeds, the unspecialised cells begin to acquire specific properties and gradually become specialised. The process by which an unspecialised cell becomes specialised (or differentiated) is known as **cell differentiation**. A few examples of differentiated cells are epithelial cells, RBCs, WBCs, cardiac cells, neurons and muscle cells. It is important

© NCERT not to be republished

to note that not all of our body cells are differentiated. Some populations of unspecialised/undifferentiated cells, also known as **stem cells**, are found to be present in organisms even after attaining adulthood. Stem cells are unspecialised cells which divide to produce two daughter cells, one of which remains a stem cell and the other one becomes differentiated. The major sources of stem cells are embryos and adult tissues (adult stem cells). The embryonic stem cells can be derived from embryos which are in the early phase of development, usually a week old blastocyst is preferred. It is a common practice to extract the inner cell mass from the blastocyst and culture the cells in a culture dish containing essential nutrients. In the total absence of any necessary stimulation to differentiate, these cells begin to divide while retaining the ability to differentiate in any type of cell. The adult stem cells, on the other hand, can be found in various tissues throughout the body, e.g., bone marrow, blood vessels, brain, skeletal muscles and liver. In these tissues, the adult stem cells remain in a non-dividing (or quiescent) state until they are finally activated due to tissue injury.

Cellular differentiation is controlled by major molecular processes involving signalling. The principal signalling molecules which aid in communication from one cell to another in the process of differentiation are known as growth factors. How this signal transduction is carried out in cells shall be discussed later in this chapter. The ability of stem cells to differentiate into other cell types is known as **cell potency**. Since stem cells can give rise to any type of cell, they are considered to be the most potent. We can classify them on the basis of their potential to differentiate into 3 types:

1. *Totipotent Cells*

A totipotent cell is able to differentiate into all possible cell types in an organism. This ability of the cell is referred to as totipotency. Totipotent cells, thus, exhibit highest differentiation potential. Unlike pluripotent cells, totipotent cells can give rise to embryonic as well as extraembryonic cells. Examples of totipotent cells include a zygote formed after fertilisation (Fig 5.19), and asexual spores.

© NCERT not to be republished

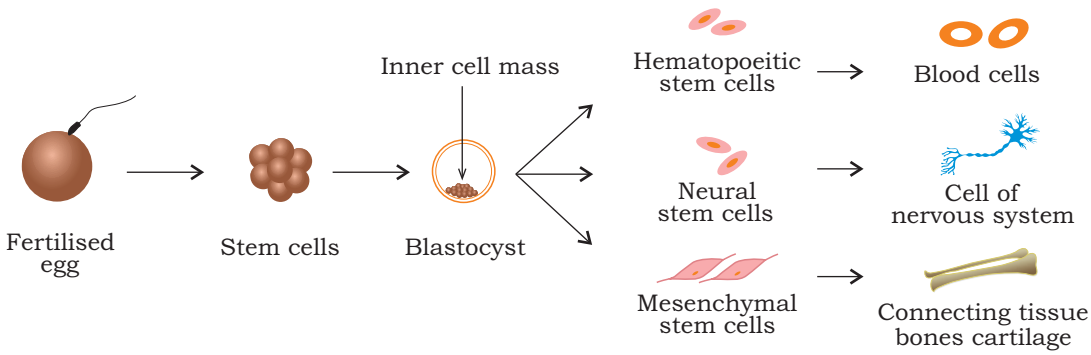


Fig. 5.19: Formation of different kinds of specialised cells from a fertilised egg (totipotent)

2. Pluripotent Cells

These cells do not exhibit full potency in the sense that they can get differentiated into most tissues of the body excepting extraembryonic tissues such as placenta or yolk sacs. Their differentiation potential is, thus, less than that of totipotent cells. An example of pluripotent stem cells include the cells derived from inner cell mass. They essentially give rise to three germ layers namely, endoderm, mesoderm and ectoderm. These can further differentiate into most of the body tissues and organs.

3. Multipotent Cells

These cells have a limited range of tissues into which they can differentiate. Therefore, they have an even lower cell differentiation potential. An excellent example of this cell type is the haematopoetic stem cells (HSCs). HSCs differentiate into a large number of blood cells including red blood cells (RBCs), white blood cells (WBCs) and platelets.

Box 1

Induced Pluripotent Stem Cells (iPSC)

Pluripotent cells are of great importance to scientists as they allow for modelling human diseases, such as cancer, congenital heart disease, etc. Embryonic stem cells (ESCs) were the standard source of pluripotent stem cells until now. However, because of ethical constraints related to death of embryos, this method is no longer in use. A new technology called Induced Pluripotent Stem Cell (iPSC) technology was developed in 2006 by Shinya Yamanaka, Japan, which uses adult tissue cells to induce pluripotency in them via expression of specific transcription factors. This bypasses the need for obtaining embryos. iPSCs can be used to make patient-matched pluripotent cells. This avoids the risk of transplant rejection.

© NCERT not to be republished

5.6 Cell Migration

Cell migration refers to the movement of cell(s) from one place to another. It occurs in unicellular organisms like amoeba to multicellular organisms like mammals and in different environments (natural soil to experimental set-ups Petri dishes). Cell migration occurs in an organism under different conditions like, during embryogenesis to create new layers, organogenesis and regeneration to generate different organs, feeding requirements and immune responses.

Depending on the type of cell and the various conditions, there are different modes of the cell migration. Cells can move as single unit or in groups. The cell migration varies with different intrinsic factors of a cell(s) like— the organisation of cytoskeleton (if it is highly organised or not), the extracellular matrix, adhesion strength and the migratory signals, etc.

Cell migration involves several stages of highly coordinated and integrated signalling networks of events. These processes comprise of polarisation (spatial differences in shape, structure, and function within a cell), protrusion and adhesion, translocation of the cell body and retraction of the rear (Fig. 5.20).

(i) Foremost, cells develop polarisation by means of specialisation of the plasma membrane/cell cortex. The polarity is reinforced and often even arises from environments that offer a directional signal. These directional signals can be chemical-based, electric field-based, mechanical force-based or concentrations of substrate-based. The end result is a definite cell front and a rear. During cell migration, the branching and polymerisation of actin filaments facilitate the extension of the leading edge.

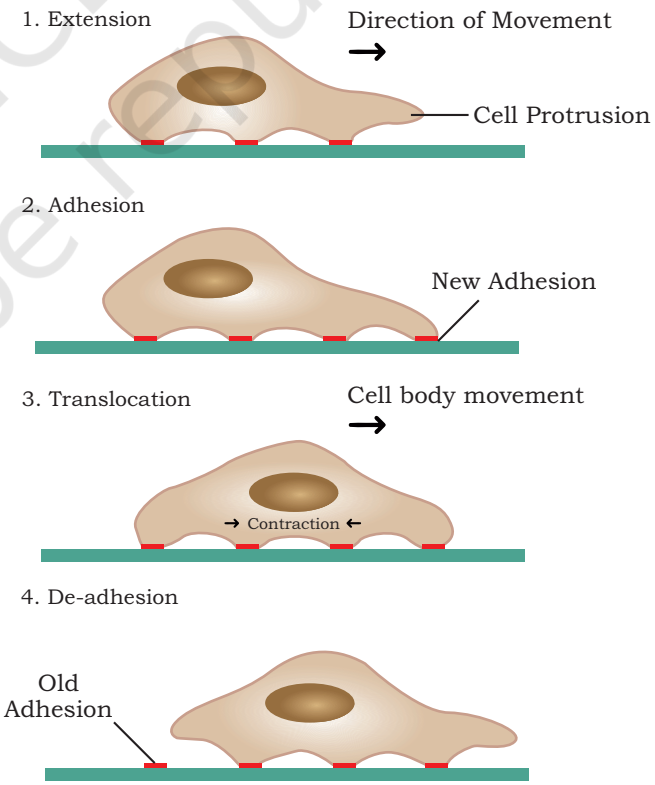


Fig. 5.20: Diagrammatic representation of the steps involved in cell migration

© NCERT not to be republished

(ii) To create a leading edge of the cell, protrusions which are membrane extensions in the direction of cell migration like pseudopodia etc., are required to be formed and extended. This comprises of the three chief components; the spreading out of the plasma membrane, the development of a core backbone that hold the expanded plasma membrane and the setting up of contact with the substratum, which provides foothold for the migration of the cell body and signals or cues that control actin polymerisation.

(iii) These protrusions attach to the substratum across which the cell is moving. Eventually, the lagging edge of the cell then separates from the substratum and is pulled back in the cell body.

(iv) Adhesions are sites of molecular communication between the cell and the substrate. They assemble and disassemble in reaction to extracellular signals and regulate cell migration. During cell migration, at the leading edge adhesions assemble and disassemble at the trailing edge. However, in protruding areas of cells, as new adhesions form, they can dissemble or stabilise and develop into more fully grown, well-built adhesions. Hence, the extremely motile cells, particularly those *in vivo*, do not have the well-built adhesions that characterise less motile cells.

(v) For cells to move forward, adhesions at the rear end are required to disassemble and the trailing edge retract, or else under tension the cell will be ripped apart. For this, the termination of the adhesive contacts at the rear edge is attained through down regulation through phosphorylation by phosphatases, or merely by proteolysis of adhesive contacts by proteases enzymes.

Role of Cell Migration

(i) Cell migration plays a fundamental role in the origin and development of morphological characteristics of embryos during **gastrulation**, the formation of the embryonic layers in the embryo by migration of large number of cells (endoderm, mesoderm and ectoderm), as well as the **organogenesis**, formation of organs and tissues. These migratory cells reach their ultimate destinations and undergo differentiation and produce the different organs and limbs.

© NCERT not to be republished

(ii) Cell migration is a key element for maintaining the **homeostasis**, the ability to maintain a steady internal environment in response to environmental changes of an individual. For example, the tissue regeneration/ repair and inflammation (during injury or infection) are important events of homeostasis. Inflammation involves the migration of immune cells from the lymph nodes to the circulation, where they remain alert until an inflammatory reaction is triggered which leads them to respond to injury or infection. Therefore, failure of cells to migrate can cause severe life-threatening defects, for example, autoimmune diseases, defective wound repair, etc.

(iii) Cell migration also plays an important role in **metastasis** i.e., the spread of cancer from one place to another. Due to loss of cell-cell interactions and increased cell motility, the cancer cells develop invasive properties and migrate and spread from the primary site of tumor growth.

© NCERT not to be republished

SUMMARY

- All cells respond to signals in their environment through specific receptors which can either be located on the cell surface or present inside the cytoplasm or the nucleus. A ligand is a chemical messenger which is specific to a receptor. Once bound to its ligand, the receptor undergoes conformational changes, allowing the signal to be interpreted. There are three types of signaling- Paracrine signaling, Autocrine signaling and Endocrine signaling.
- Living organisms take up and utilise free energy through metabolism.
- Glucose acts as an important metabolic fuel in animal. The conversion of glucose into pyruvate occurs through glycolysis.
- Under aerobic conditions, pyruvate enters the citric acid cycle in the mitochondrial matrix and gets converted into acetyl CoA. The final products of citric acid cycle is CO_2 and H_2O.
- Oxidation of fatty acids to acetyl CoA occurs through the β-oxidation pathway.
- Transamination of amino acids involves transfer of the

amino nitrogen from an amino acid to a carbon skeleton to form other amino acids. Deamination involves excretion of the amino nitrogen as urea.

- Photosynthesis is the process by which light energy gets converted into chemical energy. Photosynthesis occurs inside the chloroplast in plants and some photosynthetic bacteria. During the light reactions, solar energy is captured and splitting of water occurs along with the release of oxygen. Dark reactions (or calvin cycle) involve assimilation of carbon into trioses and hexoses, leading to synthesis of sugar.
- Cell division gives rise to two daughter cells which carry the same genetic makeup as the parent cell. Interphase and mitotic phase are important phases of cell cycle. Interphase is further divided into G1 phase, S phase and G2 phase. During interphase, the cell grow and replicate their genetic content. M phase in mitosis consists of four stages: prophase, metaphase, anaphase and telophase.
- Unlike mitosis, meiosis has a single interphase which is followed by two nuclear divisions named as Meiosis I (reductional division) and Meiosis II (equational division). M-phase is followed by cytokinesis (division of cytoplasm).
- Apoptosis refers to programmed cell death which occurs as a part of an organism's growth. It is a normal and highly regulated process.
- The process by which an unspecialised cell achieves a specialised state is referred to as cell differentiation. Stem cells are undifferentiated and unspecialised cells. The ability of stem cells to become differentiated into other cell types is known as cell potency. Stem cells are characterised into three types depending on their potency: totipotent cells, pluripotent cells and multipotent cells.
- Cell migration is the movement of cell from one place to another. It is important at various stages of development of a living organism, such as embryogenesis, organogenesis, regeneration, immune responses, etc.

© NCERT not to be republished

EXERCISES

1. Give a comparative account of the following:
   (a) Apoptosis and necrosis
   (b) Autocrine and paracrine signaling
   (c) Anabolic and catabolic pathways
   (d) Totipotent and pluripotent cells
2. Explain how stem cells are different from blood cells in terms of potency?
3. How many mitotic divisions will produce 64 cells out of a single cell?
4. Match the following

(a)	Centromere	(i)	Reductional division
(b)	Kinetochore	(ii)	Holds the two sister chromatids together
(c)	Metaphase	(iii)	Pairing of homologous chromosomes
(d)	Zygotene	(iv)	Equational division
(e)	Pachytene	(v)	Assembly of homologous chromosomes on metaphase plate
(f)	Meiosis I	(vi)	Site of attachment of chromosomes to spindle fibers
(g)	Meiosis II	(vii)	Crossing over between homologous chromosomes occurs

5. How is cell migration important for the development of an embryo?
6. What is the significance of citric acid cycle?
7. The first reaction in glycolysis is catalysed by which of the following enzymes:
   (a) Glucokinase
   (b) Phosphoglycerate kinase
   (c) Phosphofructokinase
   (d) Hexokinase

© NCERT not to be republished

8. Which of the following statements regarding cell signaling is NOT true?
   (a) In endocrine signaling, the ligand reaches the target cell through bloodstream after being synthesised in the extracellular space.
   (b) Synaptic signaling is an example of paracrine signaling.
   (c) Ligands bind to receptors in a non-specific manner.
   (d) Cancer cells exhibit autocrine signaling.
9. Dark reaction in photosynthesis is named so because—
   (a) It occurs independent of light energy
   (b) It occurs in dark
   (c) It cannot occur in daylight
   (d) All of the above
10. Differentiate between cyclic and non-cyclic photophosphorylation.
11. Briefly describe Calvin cycle.

© NCERT not to be republished

Chapter 6
Basic Principles of Inheritance

Chapter 7
Basic Processes

Chapter 8
Genetic Disorders

Unit III
Genetic Principles and Molecular Processes

The idea of inheritance patterns emerged from the work of Mendel and other scientists who followed him. What was not clear was the nature of the 'factors' which are responsible for determining a particular phenotype. It became crucial to have an understanding of the structure of genetic material and the patterns of inheritance. The foundation of molecular biology and genetics was laid down by many eminent scientists of that time, such as Watson, Crick, Nirenberg, Khorana, Monod, Benzer, etc. The contributions of these scientists and the concepts explained by them have been discussed in three chapters of this unit.

© NCERT not to be republished

Gregor Johann Mendel (1822-1884)

Gregor Johann Mendel was born on July 22, 1822 in Austria. His pioneering work laid the foundation of science of genetics and therefore, he is known as 'father of genetics'. In 1843, Mendel began studying even while being a monk at St. Thomas Monastery in Brno. There he was exposed to the lab facilities and got interested in research and teaching. His experiments focussed on cross-breeding of pea plants and gathering data on the variations for several generations. Based on his experiments on a total of seven characteristics in garden pea, he established Law of segregation and Law of independent assortment. Decades after his death in 1884, his work got recognition by other researchers. His research is now considered to be the basis of modern genetics.

© NCERT not to be republished

11150CH06

Chapter 6

Basic Principles of Inheritance

6.1 *Introduction to Inheritance*
6.2 *Linkage and Crossing Over*
6.3 *Recombination*
6.4 *Sex-linked Inheritance*
6.5 *Extrachromosomal Inheritance*
6.6 *Polyploidy*
6.7 *Reverse Genetics*

6.1 Introduction to Inheritance

Have you ever noticed that all the members of your family have several features in common like facial features, hair colour, skin colour, etc.? Why is it so? Why do you resemble in certain characters with your mother and certain characters with your father? Characteristics that run in families have a **genetic basis**, meaning that they depend on genetic information a person inherits from his or her parents. The same is true for all plants and animals.

Gregor Johann Mendel (1822–1884), 'father of genetics'

This transmission of characters from one generation to the next, or the phenomenon of the offsprings to inherit the parental traits is known as '**Heredity**'. The inherited characters are present on the chromosomes in the form of genes. Further, it is observed that though offsprings inherit

© NCERT not to be republished

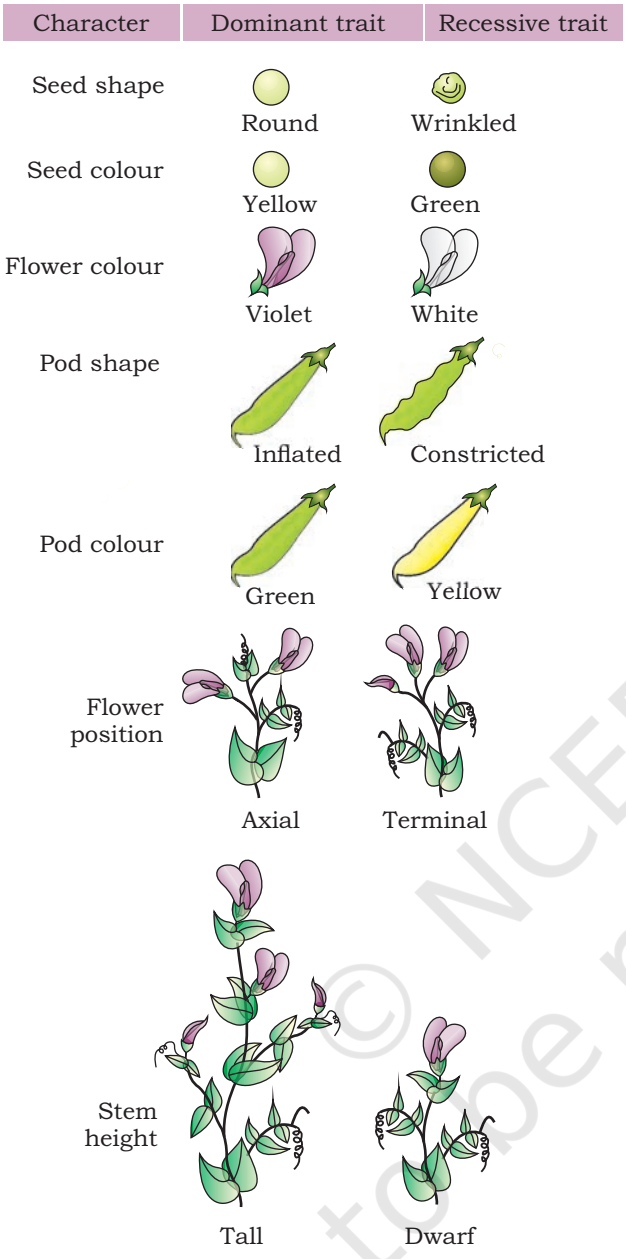


Fig. 6.1: Seven pairs of contrasting traits of pea plants used by Mendel

characters from their parents, they are unique and differ from their parents in certain aspects. These differences between the offsprings and their parents are known as **variations**. The study of scientific facts of heredity and variation is referred to as **Genetics**.

The major objective of biotechnology is the manipulation of the living organisms or to modify the genetic constitution of an organism to manufacture products intended to improve the quality of human life. In order to use biotechnological tools for manipulating the genes, understanding of the genetics and heredity of the traits is essential. It is essential to identify the genetic constituents (genes and their allelic forms in the population) regulating a trait, for its manipulation. In this chapter we will study about the principles of inheritance.

6.1.1 Mendel's work: The foundation

Our modern understanding of inheritance of traits through generations comes from the studies made by Gregor Mendel, an Austrian monk. He selected pea plants (*Pisum sativum*) for his breeding experiments as a good model system because it is an annual plant with perfect bisexual flowers and having many contrasting pair of characters. He selected seven pairs of contrasting characters for his breeding experiments and produced pure line for each trait by self-pollinating for several generations (Fig. 6.1). He performed artificial cross pollination in plants with contrasting traits by transferring pollen from one flower to another with a small brush. He grew a large number of plants for each cross and collected data for several generations.

© NCERT not to be republished

Single gene inheritance

When Mendel cross pollinated a pure (homozygous) tall pea plant with a pure dwarf pea plant, he noticed that the progeny of first generation (First filial or F_1 generation, which was raised by collecting the seeds produced from this cross) were all tall. The dwarf phenotype was missing. What happened to the dwarf trait? When the said F_1 offspring were self-pollinated to raise F_2 generation, surprisingly both tall and dwarf plants appeared in the ratio of 3:1 (3 tall and 1 dwarf). Since Mendel designed this experiment by considering only one contrasting trait, i.e., tall and dwarf, this cross is called **monohybrid cross** (Fig. 6.2). Interestingly, in all such monohybrid crosses involving other contrasting pair of characters carried out by Mendel, similar ratio of approximately 3:1 were obtained in F_2 generation. These results prompted Mendel to propose that each individual has two factors for each character (trait) and that one factor (which was later named as gene) was inherited from each parent through gametes.

This is the reason that the dwarf feature which was not there in F_1 generation was found in F_2. Hence, F_1 tall plants are heterozygotes as they contain two different alleles (Tt). As F_1 plants are heterozygous tall (Tt), this indicates that the tall allele (T) is dominant over dwarf allele (t). Thus, dwarf allele (t) is recessive to tall allele (T).

Understanding of these crosses can be well understood by the graphical representation developed by Reginald C. Punnett, a British geneticist. Using Punnett Square, we can easily calculate the probability of all possible genetic

Rediscovery of Mendel's Work

Mendel carried out hybridisation experiments on pea plants for nine long years and published all his observations in 1866 in Annual Proceedings of Natural History Society of Brünn, demonstrating the actions of invisible 'factors' now called gene, in predictably determining the traits of an organism. Mendel's conclusions were largely ignored by the vast majority. In 1900, however, his work was 'rediscovered' by three European scientists, Hugo de Vries, Carl Correns, and Erich von Tschermak.

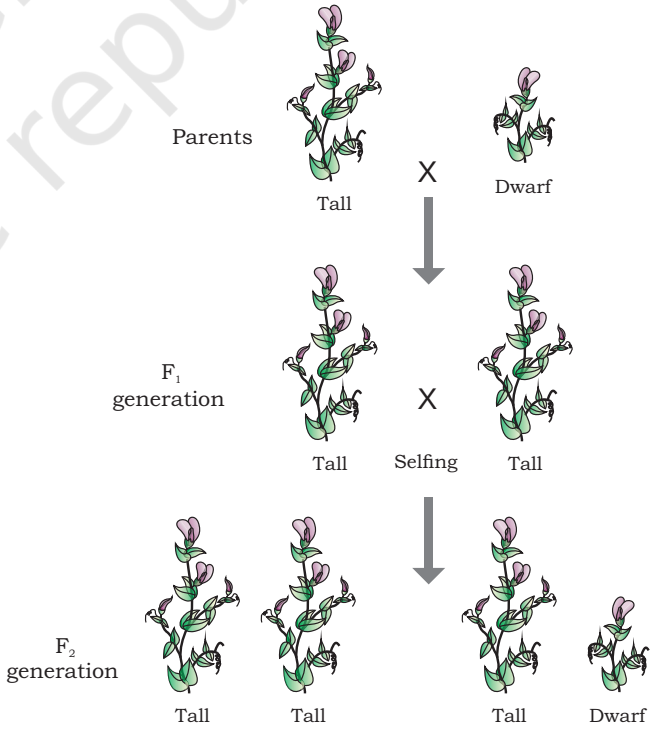


Fig. 6.2: Monohybrid cross

© NCERT not to be republished

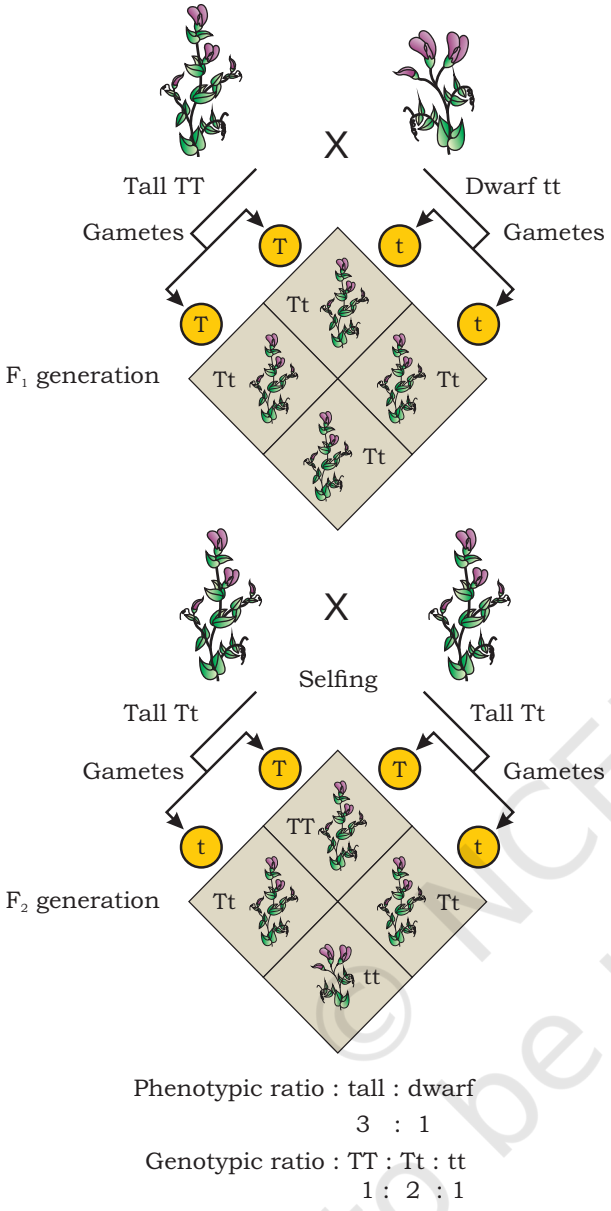


Fig. 6.3: Segregation of height character in pea plant

combinations or genotypes. We can see in the Fig. 6.3, that when plants in F_1 heterozygous progeny were self-pollinated as they produced 'T' and 't' gametes, the progeny revealed three genotype combinations; TT, Tt, tt in a ratio of 1:2:1 respectively. Here we learnt that through Punnett Square by using mathematics, we can easily calculate probability of genotype (genetic make up) and phenotype (morphological or observable traits) of future progeny. This clearly shows that the phenotypic ratio of a monohybrid cross is 3:1 and the genotypic ratio is 1:2:1. Are you able to tell about the genotype of a particular plant merely by looking at it? For example, can you say that the tall plant of F_1 or F_2 progeny has the genotype TT or Tt? For such an analysis, the F1 and F2 tall plants are crossed with dwarf plants. The appearance of tall and dwarf plants in the succeeding generation confirms the heterozygous genotype of the said plant. Such a cross is called a test cross.

Inferences can be drawn that out of two contrasting characters, one is dominant and the other is recessive. This is what Mendel's law of dominance is all about. Also, alleles of these traits do segregate while getting inherited as we have seen in the above cross, called **law of segregation**.

Incomplete dominance

When similar experiments were conducted with other pea varieties, it was observed that F_1 hybrids were not related to either of the parents but exhibited a blending/intermediate of characters of the two parents. It means the two alleles of one trait are not related as dominant and recessive, but the dominant gene in heterozygous condition has reduced expression, so that each of the allele gets expressed itself partially, called **incomplete dominance**. In four-o'clock plant, *Mirabilis jalapa*, when homozygous plants with red flowers (RR) are crossed with the homozygous plants having white

© NCERT not to be republished

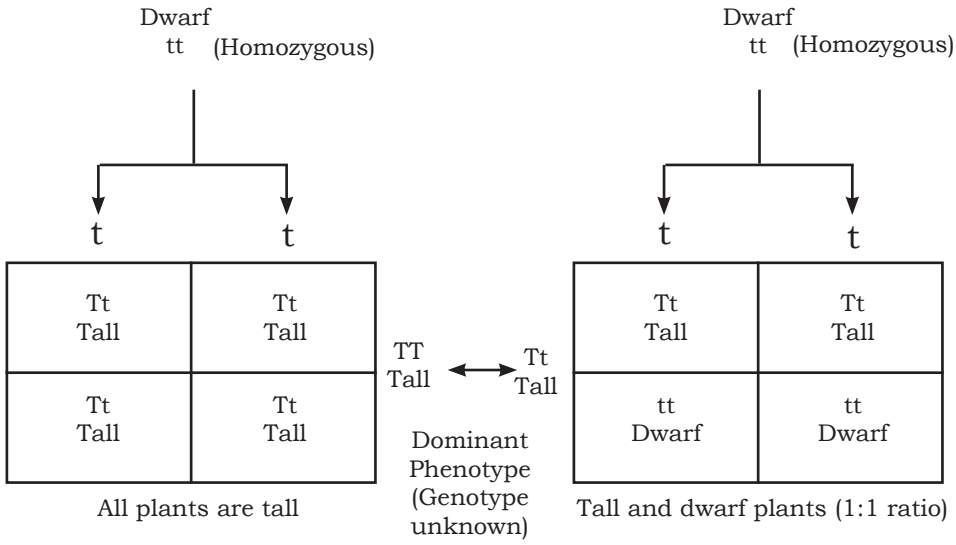


Fig. 6.4: Test cross for identification of genotype

flowers (rr), the F_1 plants (Rr) bear pink flowers, when these F_1 plants with pink flowers undergo selfing, they yield 1:2:1 ratio of red, pink and white (Fig. 6.5).

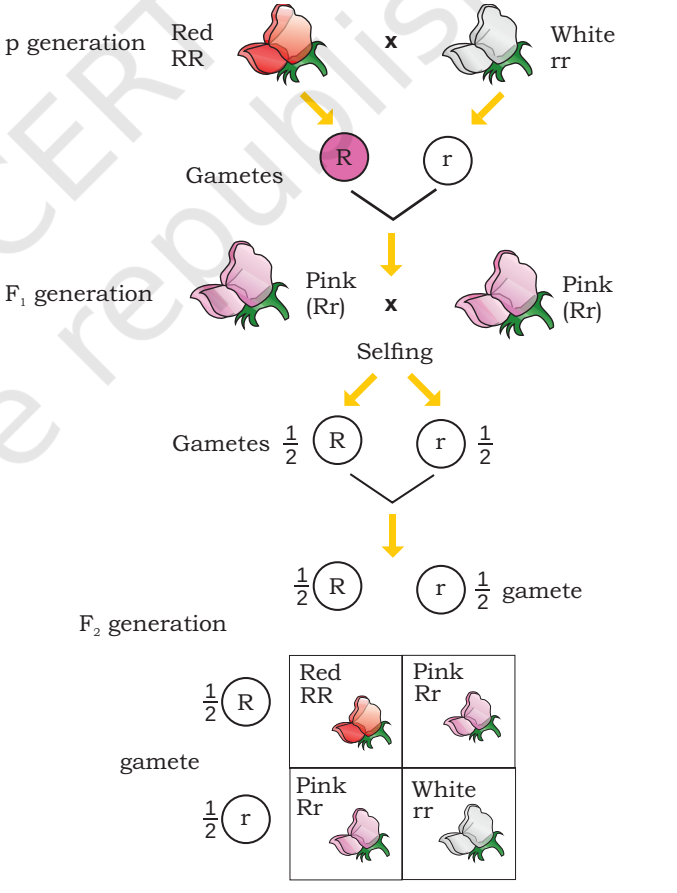


Fig. 6.5: Incomplete dominance in four-o' clock plant

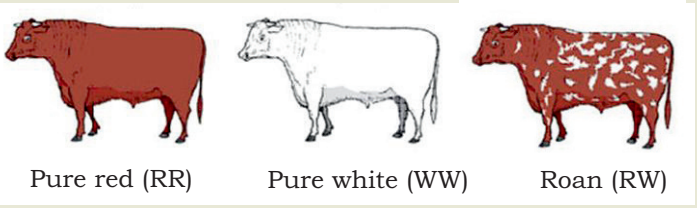
Codominance

So far we have seen that both alleles in heterozygous condition have dominant recessive relationship expressing only the dominant trait or having incomplete dominant relationship producing an intermediate trait. Many instances have been seen in which alleles of both parents get equally expressed in F_1 heterozygote. This condition is known as **codominance**. This is observed in coat colour of cattle or MN blood group of human beings (Fig. 6.6). Inheritance of coat colour in many cattles such as horses, cows and dogs is an example of codominance. When pure red (RR) is crossed with pure white (WW), the F_1 generation will have Roan (RW) coat colour, which is a heterozygous. Roan coat colour is a mixture of white and

© NCERT not to be republished

pigmented coat colours that does not fade as the animal ages. Both the red (RR) and white (WW) traits are equally expressed in F_1. Therefore, F_1 generation progeny will have roan coat colour.

Genotype	Phenotype	Antigen present on red blood cell
L^ML^M	M	M
L^ML^N	MN	M and N
L^NL^N	N	N



Pure red (RR) Pure white (WW) Roan (RW)

Fig. 6.6: Codominance of MN blood group and coat colour in cattle

Law of independent assortment

Let us now consider a dihybrid cross between homozygous round shape and yellow colour (RRYY) seeded pea plant with a homozygous wrinkled and green colour (rryy) seeded pea plant. All F_1 progeny were round seeded having yellow colour. Can you guess in this example which traits are dominant and which are recessive? In F_1 progeny, as all plants were round and yellow seeded, it clearly showed that they are dominant over wrinkled and green seeded traits.

The result of F_2 generation upon selfing is explained in Fig. 6.7 in which a ratio of 9:3:3:1 of offspring with 9 round yellow, 3 wrinkled yellow, 3 round green, and 1 wrinkled green (9:3:3:1) is observed. Since two pairs of contrasting characters are included in such crosses, hence they are called **dihybrid crosses**.

Based upon such observations on dihybrid crosses, the third principle of inheritance, i.e., **Law of Independent Assortment** was proposed.

There is an interesting observation in such a dihybrid cross that not only the parental traits reappear in F_2 but there are new combinations of traits, i.e., round shaped seed with green colour and wrinkled seed with yellow colour (Fig. 6.7). Such a new combination is possible only in a situation when factors or genes controlling a specific trait are inherited independent of each other. Such a pattern of inheritance is known as the principle of independent assortment of alleles. *Can you work out the genotypic ratio of F_2 progeny using the Punnett square data?*

© NCERT not to be republished

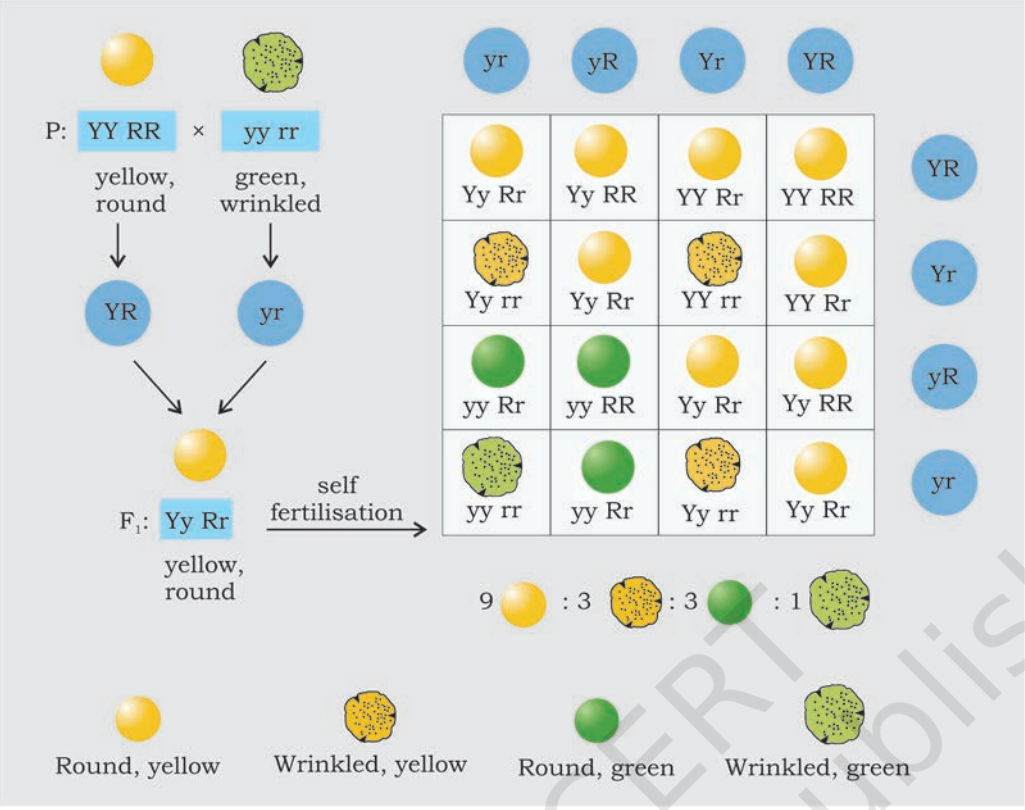


Fig. 6.7: Results of a dihybrid cross where parents differ in two pairs of contrasting characters

6.2 Linkage and Crossing Over

We have already learnt that there are several phenotypic traits in the body of an organism such as colour of flower (red/white), shape of pollen (round/elliptical), etc., in pea. Each of these phenotypic traits is determined by a pair of alleles, which is located at a specific gene loci of homologous chromosomes (autosomes or sex-chromosomes). Thus, organisms may have numerous genes for its various phenotypic traits. As you know in humans, there are 20,000 to 25,000 protein coding genes present on 23 pairs of chromosomes. Thus, each chromosome contains several genes. Can you think the genes present in each chromosome are inherited together or independently? Because several genes are present in a chromosome, they should be inherited together as a unit during meiosis. This phenomenon of the inheritance of genes together and to retain their parental combination even in the offspring is known as **linkage**. The genes located on the same chromosomes and being inherited together are known as

© NCERT not to be republished

linked genes and the characters controlled by these genes are known as **linked characters**. All the genes located on the single chromosomes constitute a linkage group.

W. Bateson and R.C. Punnett provided the evidence in favour of linkage in their experiments on sweet pea (Fig. 6.8). They crossed plants with red flowers and long pollen grains to plants with white flowers and short pollen grains. All plants in F_1 progeny/generation had red flower with long pollen grains, thus indicating that the alleles for these two phenotypes were dominant. When the F_1 progeny was self

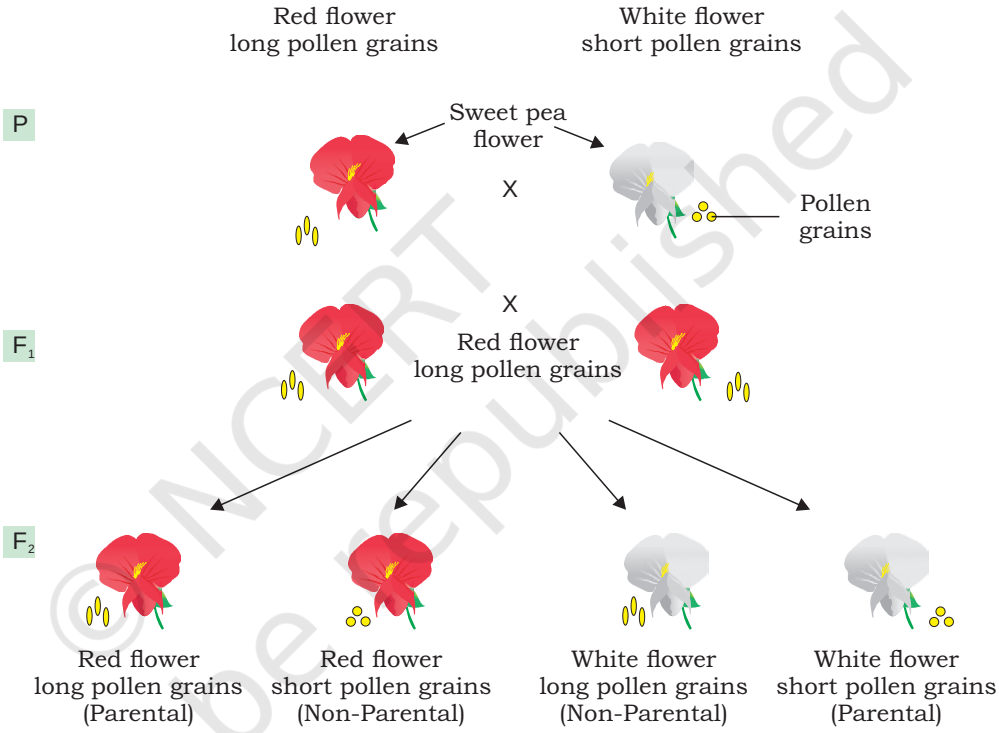


Fig. 6.8: Bateson and Punnett experiment on sweet pea to study linkage

pollinated, they observed peculiar distribution of *genotypes* among the offspring (Fig. 6.8). Bateson and Punnett could not provide the correct explanation for this experiment but later on in similar kind of experiments on *Drosophila* performed by Morgan and his colleagues in 1910 provided explanation for the same which is discussed in the next section.

The data revealed that the genes for flower colour [Red (R), white (r)] and pollen grain length [Long (L), short (l)] do not assort independently as expected. The correct explanation for the lack of independent assortment in the

© NCERT not to be republished

data is that the genes for flower colour and pollen length are located on the same chromosome, that is they are linked. This is explained in the diagram (Fig. 6.9).

Later on Morgan (1910) suggested that the genes are present in a linear fashion in chromosome. All the genes present in the same chromosome are inherited together generation after generation retaining the parental combination. A cross between a homozygous grey bodied vestigial winged (BBvv) *Drosophila* with a black bodied long winged (bbVV) *Drosophila* produced grey bodied and long winged (BbVv) flies in F_1 generation. When these flies were crossed with a double recessive fly (bbvv), surprisingly in addition to parental combination (83%), non-parental (17%) combination appeared. This indicated that linked genes do not stay together always but may get separated due to exchange of segments during gametogenesis. This phenomenon of interchange of chromosome segments is known as **crossing over** (Fig. 6.10). The linked genes are located on the same chromosome in a linear fashion. If chromosomes remain intact during inheritance, the genes located on one chromosome should be inherited together generation after generation, and only parental combinations must appear in F_2 generation. But in most of the cases though parental combinations are more numerous, non-parental combinations also appear. This indicates that the linked genes do not always stay together but get separated many a times. They get separated with an interchange of alleles, resulting in the appearance of non-parental combinations. When Morgan mated grey bodied vestigial winged (BBvv) and black bodied long winged (bbVV) *Drosophila*, it produced F_1 hybrid, all of them having grey body and long wings (BbVv). When female flies of F_1 generation were crossed

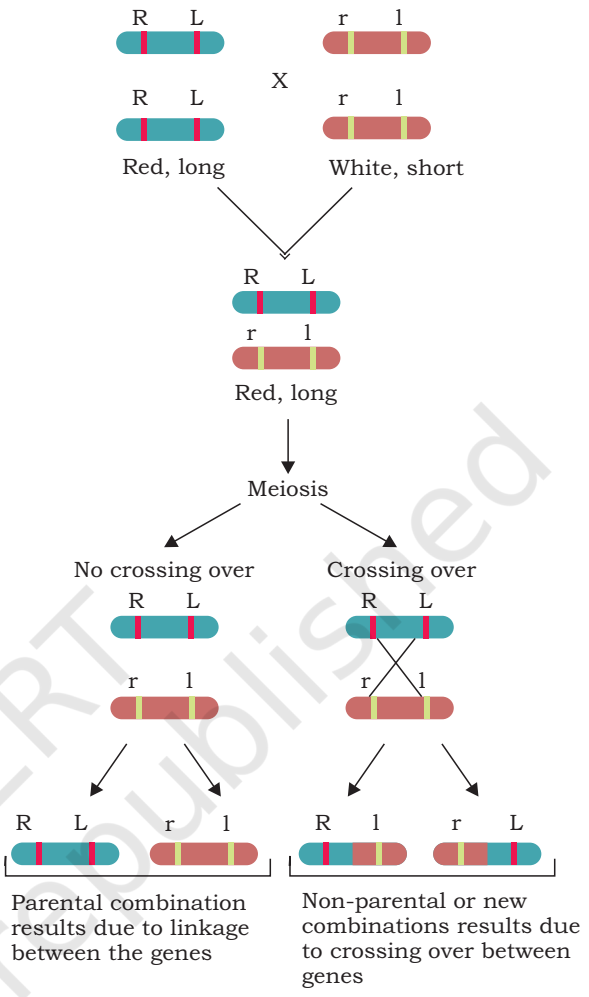


Fig. 6.9: Linkage and crossing over among genes for flower colour (R and r) and pollen shape (L and l)

© NCERT not to be republished

with double recessive males having black body and vestigial wings (bbvv), four types of offsprings were produced as follows:

Grey vestigial — 41.5 per cent
Grey long — 8.5 per cent
Black vestigial — 8.5 per cent
Black long — 41.5 per cent

In this case parental combinations are 83 per cent and non-parental combinations are 17 per cent. This phenomenon in which non-parental combinations appeared due to interchange of alleles is called **crossing over**.

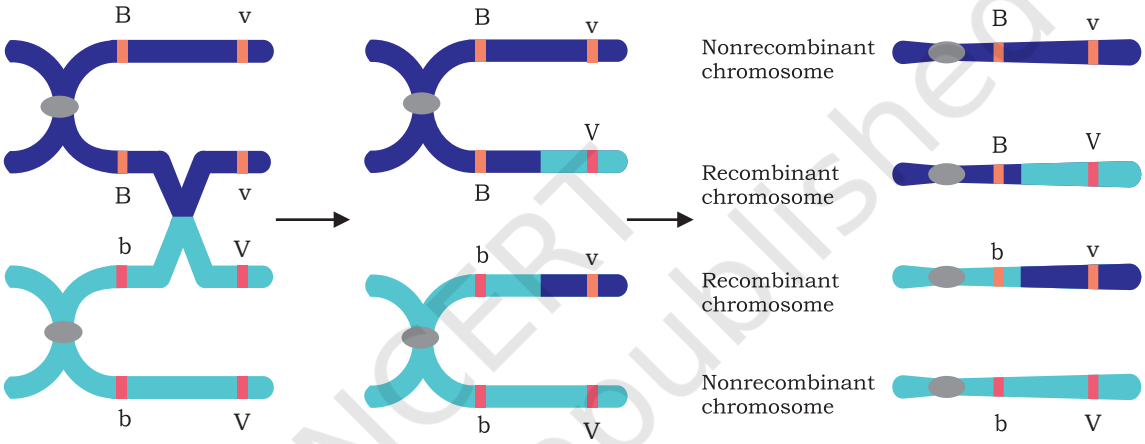


Fig. 6.10: Single cross over between two non-sister chromatids of a pair of homologous chromosomes

6.3 RECOMBINATION

The concept of recombination was very quickly observed by geneticists after the rediscovery of Mendel's principles of inheritance in the year 1900. Experiments performed by W. Bateson and R. C. Punnet in sweet pea clearly established that all genes do not assort independently. This was evident in the dihybrid cross performed in sweet pea. Crossing of pure line of red flower having long pollen grains with white flower having short pollen grains resulted into expected red flower and long pollen offspring. But among the selfed F_2 offspring more than 50% of the offspring had parental combinations of red-long and white-short for flower colour and pollen shape. Recombinants, i.e., red flower having short pollen and white flower having long pollen were less than 50%. This was a deviation from the expected 9:3:3:1 ratio (Fig. 6.8).

© NCERT not to be republished

Though, the initial explanation suggested by Bateson and Punnet were different, but later, this was established as the phenomenon of linkage in which genes situated on one chromosome inherit together. The appearance of recombinants was attributed to the phenomenon of crossing over, in which exchange of parts of homologous chromosome during meiosis may result into recombinants. Obviously, percentage of such recombinants is expected to be less in comparison to the independently assorting traits as the same has been observed in the experiment described earlier. Later, similar experiments performed in *Drosophila melanogaster* by Thomas Hunt Morgan provided many instances of linkage and recombination. Not only this, based on the frequency or percentages of recombinants, a physical map of chromosomes of a number of organisms were prepared. In preparing such physical map of the chromosome 1% of recombinants between two traits were designated to be apart by 1 map unit or 1 centimorgan (cM).

Further, the cytological studies of meiosis and appearance of chiasmata in the first meiotic division, when the homologous chromosome pair separate (reductional division) provided insight about direct connection of recombinants with the exchange of the parts of homologous chromosome pair. One of the initial evidence that crossing over is responsible for recombination came from the classical experiment performed by Harriet Creighton and Barbara McClintock

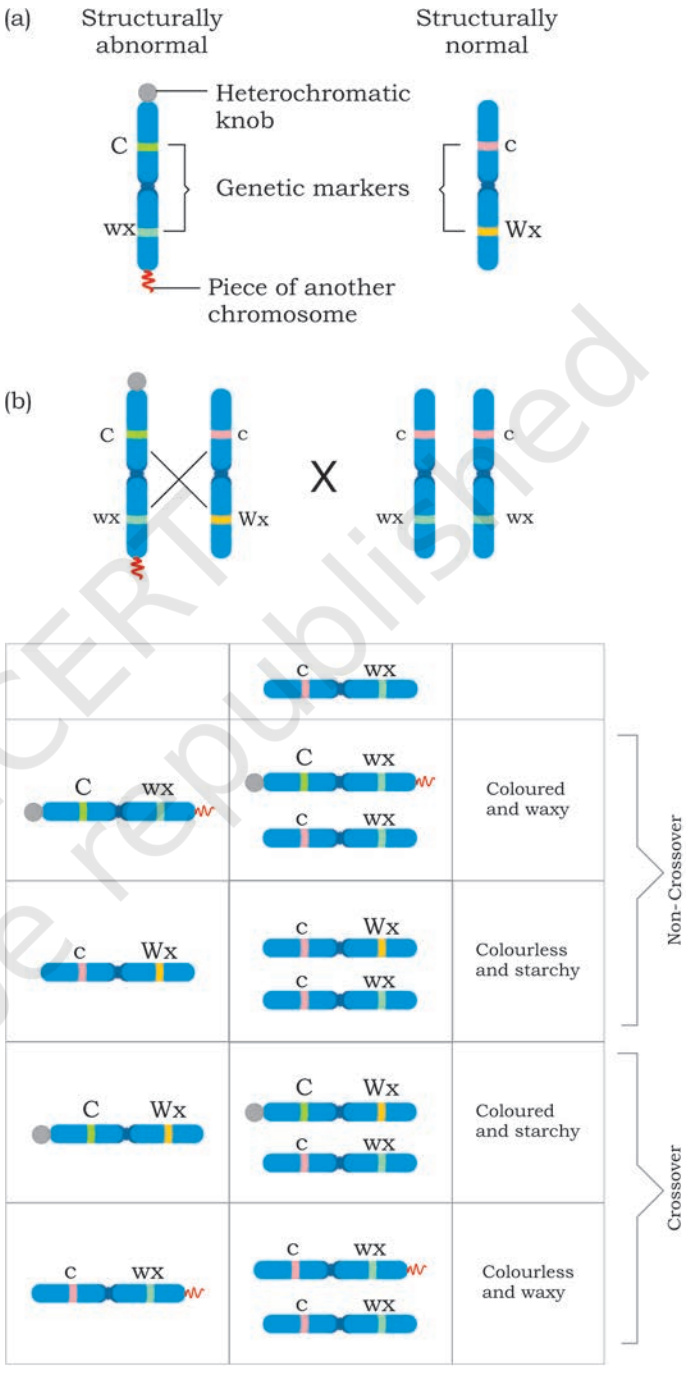


Fig. 6.11: Experimental evidence of crossing over (a) showing cytological marker on one chromosome and other normal chromosome (b) Result of test cross showing non-cross over and cross over progeny

© NCERT not to be republished

(1931) in maize. Based on their careful observation they found that some of the homologous chromosomes can be morphologically distinguished. In their investigation, two types of chromosome 9 were identified. One of the pair was a normal one whereas the other had a heterochromatic knob at one end of the chromosome and a small piece of another chromosome on the other end of it. You have already studied chromosomal aberration and understood as to how these may arise. In their experiment, two traits were used as markers to understand recombination. One was the gene that is responsible for colour of the kernel, i.e., coloured (C) versus colourless (c). The other was the gene responsible for texture of the kernel, i.e., starchy (Wx) versus waxy (wx). Performance of the cross as per Fig. 6.11 provides an evidence that during crossing over exchange of parts of homologous chromosome occur. When recombinant offspring were examined for the exchanged part of chromosome, it was found that recombinants, i.e., either coloured with starchy kernel or colourless with waxy kernel had different morphological forms of chromosome 9. Here only one of the aberrant markers were found cytologically among recombinants, contrary to the presence of both the markers on the same chromosome, establishing the fact that during recombination exchange of part of homologous chromosomes do take place.

6.4 SEX-LINKED INHERITANCE

One of the earliest known instance of sex-linked character is the bleeders disease haemophilia observed only in males in the royal family of Britain. However, the concept of sex-linked inheritance was introduced by Thomas H. Morgan in 1910, while working on *Drosophila melanogaster*.

Morgan and his coworkers noted the sudden appearance of one white-eyed male in the culture of wild red eyed *Drosophila*. This white-eyed male was crossed with red eyed female, the flies of F_1 generation (both male and female) were all red eyed, indicating that white-eyed mutation (w) is recessive to red eye colour (W). When F_1 flies mate freely, the red and white eyed flies appeared in the ratio of 3:1 in F_2 generation. But all the white eyed flies were male. The red-eyed males were equally numerous. The female on the other hand were all red eyed. The white eyed female did not appear. Morgan concluded that the

© NCERT not to be republished

gene for eye colour is located on X chromosome (Fig. 6.12). Such genes for autosomal characters present in sex chromosomes are called sex-linked genes and inheritance of these sex-linked genes is called sex-linked inheritance. Colour blindness and haemophilia are common examples of sex-linked inheritance in human.

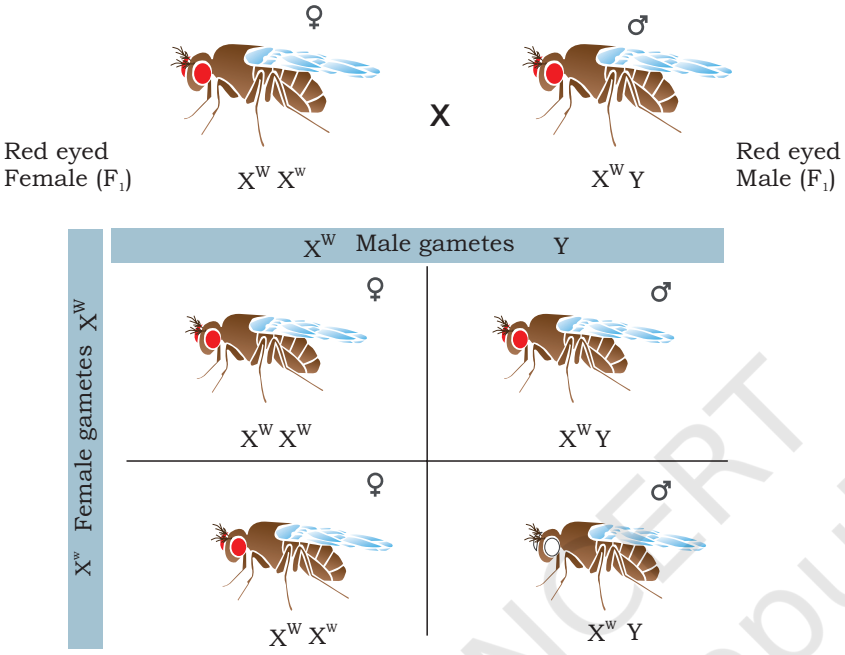


Fig. 6.12: Sex linkage in Drosophila

6.5 Extrachromosomal Inheritance

As discussed earlier, in addition to nucleus, DNA is also present in mitochondria and plastids. Gametes carry a copy of nuclear DNA from the respective parents and that combine to form new individual after fertilisation. An interesting feature of fertilisation process is that sperm cell loses most of its cytoplasm and cytoplasmic organelles before fertilisation and only sperm nucleus enters in the egg. Therefore, zygote receives genome of plastids and mitochondria only from the maternal parent. This phenomena is also called as **extrachromosomal** or **cytoplasmic inheritance**. Several traits are controlled by the genes present in plastids or mitochondrial genomes. These traits do not follow the Mendelian principles of inheritance and most of the cytoplasmic traits that have been recorded, they follow maternal line. Therefore, this phenomenon is also known as

Fig. 6.13: Plant with variegated leaves

© NCERT not to be republished

maternal or **uniparental inheritance**. For example, some of the enzymes required for cellular respiration are coded by DNA present in the mitochondria and DNA that codes for chlorophyll or other pigments is present in plastids. In Four O'clock plant (*Mirabilis jalapa*), leaves can be white, green or variegated (mixture of white and green) because of genes in the DNA of plastids (Fig. 6.13). The evidence of maternal or extrachromosomal inheritance concludingly came from the crosses conducted in the Four O'clock plant. The cross of female green, white or variegated plant with either of those make phenotypes yield offspring with female phenotype only.

6.6 POLYPLOIDY

As discussed earlier, the number of complete set of chromosomes in an organism represents the ploidy number. The organisms having one or two complete sets of chromosomes are known as haploids or diploids, respectively. Other organisms having more than two sets of chromosomes in each cell are known as **polyploids.** Depending upon the number of chromosome sets, polyploid are known as triploid (3 sets), tetraploid (4 sets), hexaploid (6 sets), octoploid (8 sets) and so on (Fig. 6.14). Most of the species that we see around us are diploid. Natural occurrence of monoploidy or haploidy is rare in nature. In some species of bees and ants, males are haploid, and females are diploid. Though, polyploidy is rare in animal kingdom, it is very common in the plant kingdom (Table 6.1). In fact, more than 30% plants are polyploid. Size of various parts like leaves and cell size in polyploid plants is typically larger as compared to diploid plants. Further, polyploid plants seem to be more tolerant to harsh environment conditions.

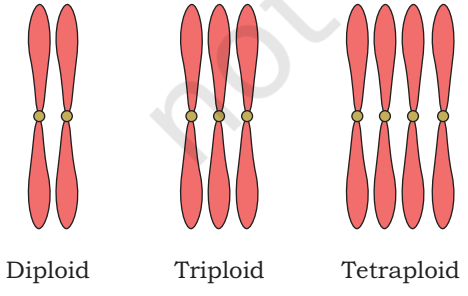


Fig. 6.14: Chromosomes in polyploid genomes

Change in number of chromosomes does not occur as complete set always. In some organisms some chromosomes may be over- or under-represented, i.e., they have incomplete set of chromosomes. These organisms in which either one or more chromosomes of a chromosome set are either missing or have more copies, are known as **aneuploids**. Aneuploidy usually results from irregular meiotic division that leads to unequal distribution of chromosomes to opposite poles.

© NCERT not to be republished

Table 6.1 List of plants representing various ploidy levels

Name of Plant	Total No. of Chromosomes	No. chromosomes in one Set (X)	Ploidy Level
Rice	24	12	Diploid
Sorghum	20	10	Diploid
Banana	22 or 33	11	Diploid or Triploid
Apple	34 or 51	17	Diploid or Triploid
Peanuts	40	10	Tetraploid
Cotton	52	13	Tetraploid
Potato	48	12	Tetraploid
Wheat	42	7	Hexaploid
Strawberry	56	7	Octoploid
Sugarcane	80	10	Octoploid

6.7 Reverse Genetics

In the above sections, we have learnt as to how we study the genetic principles based upon the phenotypic variations. These can be measured in the form of macro variations like visible morphological variations such as size, shape and number of body organs (i.e., macro-variations) or variation in DNA sequences, protein profiles, or metabolites, etc. (i.e., micro-variations). This process of analysing phenotypic variations in the population and to determine genetic constituents (DNA sequences or genes) regulating these variations is called forward genetics (Fig. 6.15). In last few decades, DNA sequencing technologies have evolved exponentially and has made it possible to read complete genome of organisms, thereby identifying all the genes in them. In a reverse genetic approach, investigation starts with the analysis of DNA or protein sequence rather than phenotypic variations. In forward genetic approach, it is possible to identify regulatory genes that produce visible phenotypes, whereas reverse genetics can be used to investigate function of any gene/protein in an organism.

© NCERT not to be republished

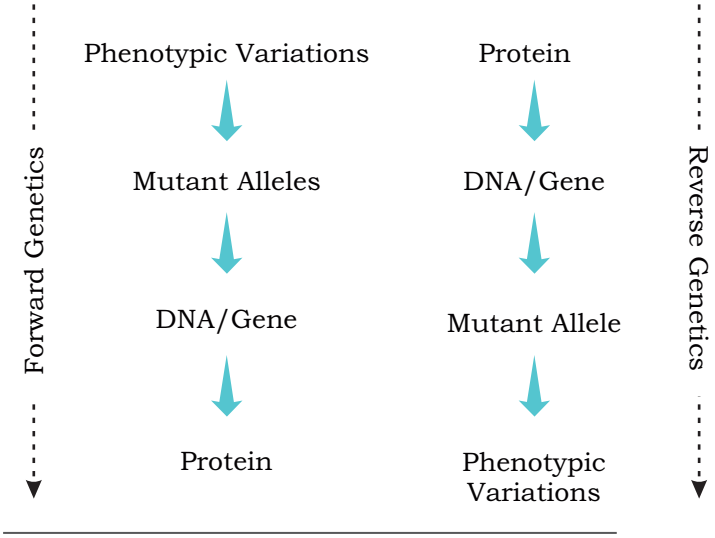


Fig. 6.15: Forward genetics & reverse genetics

Reverse genetics approach starts with DNA sequence (gene) or a *protein sequence* with an unknown function (Fig. 6.14). First candidate gene is selected whose function is not known and we want to determine its function. Various experimental procedures are used to disrupt the candidate gene and then its effect on development of the organism is analysed. If the candidate protein, the function of which is to be determined is a protein, it is first traced backwardly to ascertain its DNA sequence. The candidate gene is cloned and is reinserted back to genome of the same organism. Its expression is silenced to determine the phenotypic outcome. Whole genome sequencing has led to identification of large number of genes, whereas functions of only a few genes are known. Further, each gene in an organism does not result in visible variations. Therefore, reverse genetics approach can be used to characterise genes having unknown functions. In summary, the goal of reverse genetic approach is to induce variations in a particular gene and investigate their impact on an organism. Several technical procedures are used for causing variation (disruption or altering) in target. These techniques can be very specific like gene silencing through RNA interference (RNAi) or targeted gene disruption by homologous recombination. RNAi is a regulatory biological process, which uses small double stranded RNA molecules to inhibit gene expression or translation.

SUMMARY

- Characteristics that run in families have a genetic basis, meaning that they depend on genetic information that a person inherits from his or her parents. This is true for all plants and animals.
- Transmission of characters from one generation to the next generation or the phenomenon of the offspring to inherit

© NCERT not to be republished

the parental traits is known as Heredity.

- The differences that exist between the offsprings and their parents are known as variation.
- The study of scientific facts of heredity and variation is referred to as Genetics.
- Our modern understanding of inheritance of traits through generations comes from the studies made by Gregor Mendel, an Austrian monk.
- When a cross is made between parents with one pair of contrasting characters it is called a monohybrid cross.
- Each gene exists in two alternate forms called alleles. When two similar alleles for a trait are present in an individual it is referred as homozygous alleles, while when two different alleles for a trait are present in an individual it is referred as heterozygous alleles.
- The phenotypic ratio of a monohybrid cross is 3:1 and the genotypic ratio is 1:2:1.
- A test cross help determine whether a dominant phenotype is homozygous or heterozygous for a specific allele.
- Mendel's law of dominance states that out of the two contrasting characters, one is dominant and the other is recessive.
- According to law of segregation, alleles of traits segregate in the process of inheritance.
- In case of incomplete dominance, the two alleles of a trait are not related as dominant and recessive, but the dominant gene in heterozygous condition reduce expression, so that each of the allele expresses itself partially.
- When both the alleles of parental generation get equally expressed in F_1 heterozygote, it is known as codominance.
- When a cross is made between parents with two pairs of contrasting characters, it is called a dihybrid cross.
- According to the principle of independent assortment of alleles, factors or genes controlling a specific trait is inherited independent of each other.
- The phenomenon of inheritance of genes together and to retain their parental combination even in the offspring is known as linkage. Genes located on the same chromosome and being inherited together are known as linked genes and the characters controlled by these genes are known as linked characters.
- The phenomenon in which non-parental combinations appear due to interchange of alleles is called crossing over.

© NCERT not to be republished

- Recombination is a process during which exchange of part of homologous chromosomes take place.
- There are different types of inheritance: Sex-linked inheritance (inheritance resulting from a recessive gene in the sex chromosome), extra chromosomal or cytoplasmic inheritance (inheritance of characters controlled by genes present in the cytoplasm) and maternal or uniparental inheritance (transmission of genetic characters only from maternal extra nuclear elements such as mitochondrial DNA). Organisms having more than two sets of chromosomes in each cell are known as polyploids.
- Organisms in which either one or more chromosomes of a chromosome set are either missing or have more copies are known as aneuploids.
- The process of analysing phenotypic variations in the population and to determine genetic constituents (DNA sequences or genes) regulating these variations is called forward genetics. While reverse genetics approach starts with DNA sequence (gene) or a protein sequence with an unknown function.
- Characters are controlled by small DNA segments called genes.
- Collection of all the DNA material in one complete set of chromosomes (including all the genes and other part of DNA) in a cell is referred to as genome. A eukaryotic cell has two types of genome: nuclear genome and organellar genome (such as chloroplast genome and mitochondrial genome).
- One complete set of chromosomes in an individual makes genome of that individual and is represented by the letter n. Most of the organisms carry two sets of chromosomes (2n) and are known as diploid organisms.

EXERCISES

1. Differentiate between the following
   (a) Genotype and Phenotype
   (b) Dominant and Recessive characters
   (c) Hybrid and Pure individuals
   (d) Heterozygous and Homozygous progeny
   (e) Monohybrid and Dihybrid cross
   (f) Gene and allele
   (g) Incomplete dominance and codominance

© NCERT not to be republished

2. Mention the genotypic and phenotypic ratio of progeny when there is a cross between
   (a) F_1 progeny with pure dominant parent
   (b) F_1 progeny with pure recessive parent
   (c) F_1 progeny with F_1 progeny
3. Explain test cross through diagrammatic representation.
4. Explain following using monohybrid and dihybrid cross.
   (a) Law of dominance
   (b) Law of segregation
   (c) Law of independent assortment
5. What will be the genotypic and phenotypic ratio when a red and tall homozygous tomato plant is crossed to a red and tall heterozygous plant?
6. When one male and one female Drosophila, heterozygous for the two pairs of alleles AaBb, were mated, the offspring's phenotypic ratio 2:1:1:2 was obtained.
   (a) Explain how these ratios help in detecting linkages?
   (b) How degree of linkage can be determined?
7. Make a close observation with the nature. Do you think that the phenomenon of linkage is absolute?

© NCERT not to be republished

11150CH07

Chapter 7

Basic Processes

7.1 *DNA as the Genetic Material*
7.2 *Prokaryotic and Eukaryotic Gene Organisation*
7.3 *DNA Replication*
7.4 *Gene Expression*
7.5 *Genetic Code*
7.6 *Translation*
7.7 *Gene Mutation*
7.8 *DNA Repair*
7.9 *Regulation of Gene Expression*

7.1 DNA as the Genetic Material

You have studied in previous chapter that characters or traits are inherited from parents to offspring through genes. You are also aware that these genes are present on chromosomes which are made up of nucleic acids and proteins. However, understanding the nature of gene which is responsible for expression of trait was one of the biggest challenges before the scientific community. Answer to this question came after a few experimental evidences that deoxyribonucleic acid (DNA) determines the trait or feature of any organism except a few viruses.

Credit of discovery of DNA goes to Johann Friedrich Miescher, who for the first time isolated an acidic substance from nuclei of pus cells and named nuclein having DNA and protein. Due to presence in chromosome and nucleus these two chemical components; nucleic acid (mainly DNA) and protein became possible candidates to be the genetic material. Still, the nature of genetic material remained unknown for a long time. Gradually, experiments with microorganisms by different investigators yielded results that provided evidences in favour of DNA as genetic material.

© NCERT not to be republished

7.1.1 Discovery of the transforming principle

In 1928, a British medical officer, Frederick Griffith made an observation in the course of developing a vaccine against pneumonia caused by bacterium *Streptococcus pneumoniae* (also called *Diplococcus pneumoniae*) in mammals, which causes pneumonia in humans and is normally lethal in mice. He identified two different strains (varieties) of the bacterium i.e. virulent (disease causing) and non-virulent (harmless). In virulent strain, each bacterium is surrounded by a polysaccharide capsule because of which the bacterial colony when grown on an agar plate appear smooth and are referred to as smooth strain (S). The non-virulent strain lack polysaccharide coat and produce rough looking colony and are referred to as rough strain (R). The S type bacteria kill mice by causing pneumonia.

Griffith made a series of experiments with S and R type bacteria (Fig. 7.1). When he injected live S bacteria into mice, the mice developed pneumonia and died. However, when he infected mice with R type bacteria mice showed no ill effects. The results of these two experiments confirmed that the polysaccharide coat present in S type bacteria was apparently necessary for virulence.

In order to understand further, Griffith killed some

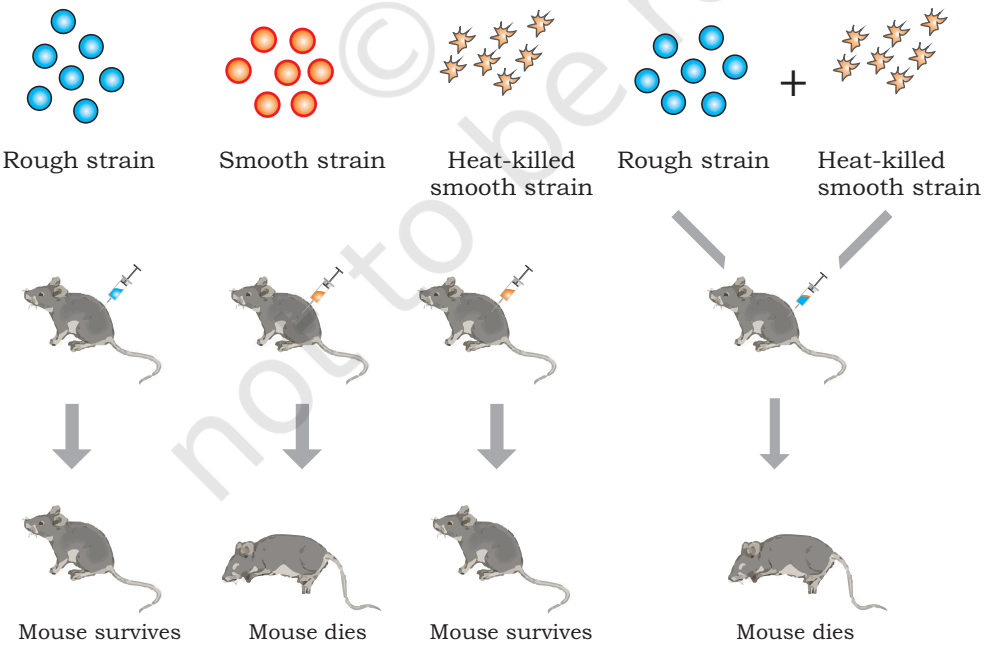


Fig. 7.1: Griffith's transformation experiment

© NCERT not to be republished

virulent S bacteria by boiling them and injected the said heat-killed bacteria into mice. As per his expectations, mice survived. However, quite unexpectedly, mice died due to pneumonia when it was injected with a mixture of heat-killed S bacteria and live R bacteria. Examination of blood and tissue fluid of the dead mice revealed the presence of live S type bacteria. Based on the above observation, Griffith concluded that the R-strain bacteria must have taken up what he called a '**transforming principle**' from the heat-killed S bacteria, which allowed them to 'transform' into smooth-coated bacteria and become virulent. He called the phenomenon as transformation, which means transfer of genetic material from one cell to another that alter the genetic makeup of the recipient cell. But the nature of transforming substance still needed to be determined.

7.1.2 Biochemical characterisation of transforming principle

Three scientists, Oswald T. Avery, Colin Macleod and Maclyn McCarty conducted a series of experiments to identify the Griffith's transforming principle, and it was confirmed in 1944 that the transforming agent is DNA (Fig. 7.2). In the design of experiment, they focused on three main components of smooth strain of bacteria, i.e., DNA, RNA and protein. They prepared an extract of heat-killed smooth strain of the bacteria from which lipids and carbohydrates were removed. Remaining components of the extract having proteins, RNA and DNA were retained for further experiment by dividing the extract into three parts. These extracts were separately treated with hydrolytic enzymes like ribonuclease (RNase), deoxyribonuclease (DNase) and protease to degrade RNA, DNA and protein, respectively, for their transforming ability by transferring each of the enzyme treated extracts into three different cultures of rough strain of bacteria. Transformation of rough strain into the smooth strain was observed in those colonies in which RNase and protease treated extract were added and not in the colony to which DNase treated extract was added. These results established beyond doubt that it is DNA which acts as a likely transforming principle.

© NCERT not to be republished

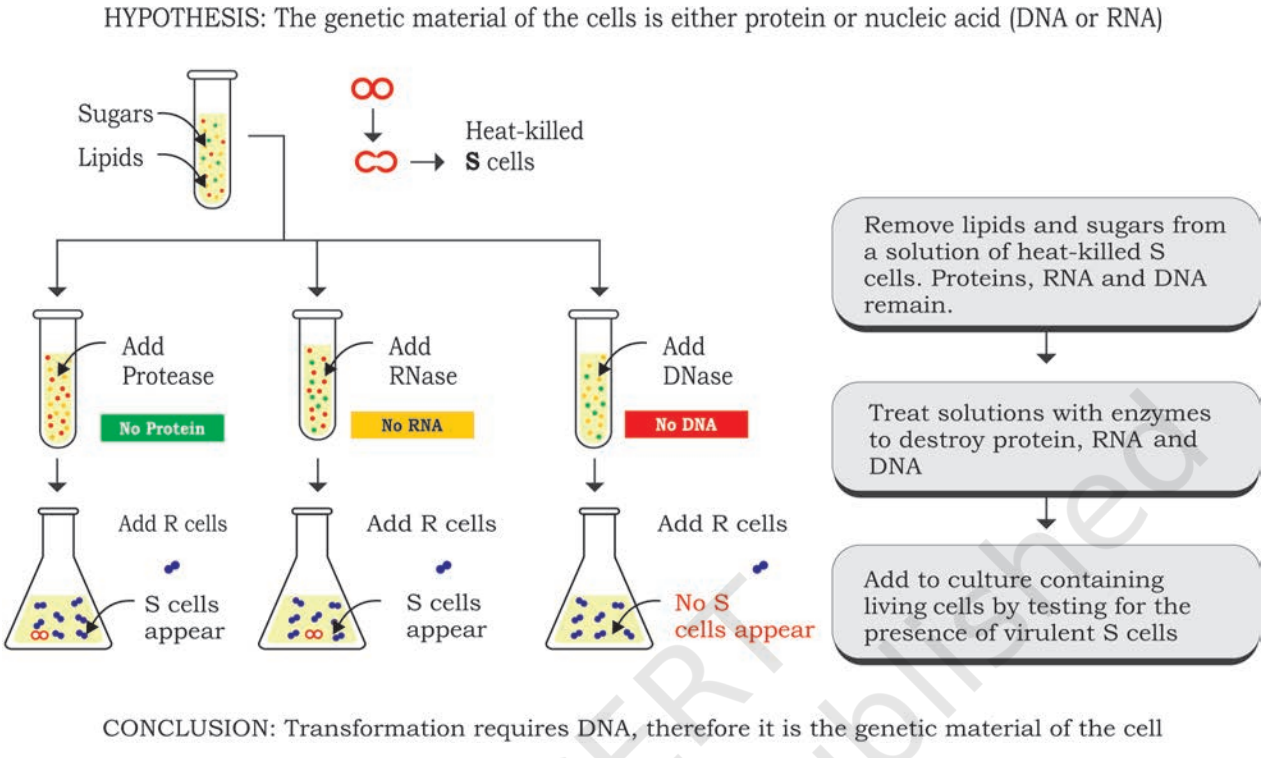


Fig. 7.2: Confirmation of transforming principle

7.1.3 The Hershey – Chase experiment

Later on, yet another experiment conducted by Alfred Hershey and Martha Chase (1952) with T2 bacteriophages provided evidence in favour of DNA as genetic material. The virus T2 bacteriophage that infects *Escherichia coli* bacteria contains DNA surrounded by a protein coat. When it infects a bacterial cell, it attaches onto the outer surface followed by injecting its DNA into the cell. In a series of their experiments with T2 bacteriophage and *E. coli*, the purpose was to establish as to which component is responsible for multiplication of phage particles, DNA or protein. To identify easily, T2 bacteriophages were initially grown with the colonies of *E. coli* in medium containing radioactive phosphorous (^{32}P) and radioactive sulfur (^{35}S) separately (Fig. 7.3). This led to labelling of one set of bacteriophages with radioactive phosphorous (^{32}P) and the other set with radioactive sulphur (^{35}S).

^{35}S and ^{32}P labelled T2 phages were now inoculated into two separate cultures of unlabelled *E. coli* bacterial colony. After infection, the bacterial colonies were agitated

© NCERT not to be republished

in a blender for removing any remaining phage and phage parts from the outside of the bacterial cells. The mixture of the blender was then centrifuged to separate the bacteria (present in pellet) from the phage debris (present in supernatant). Pellets of bacterial culture which showed radioactivity were infected with phages having radioactive DNA, whereas, radioactivity was observed in the supernatant which was infected with ^{35}S bacteriophage. This indicates that proteins did not enter the bacteria from the phage. It was therefore, concluded that the material which enters into bacterial cell, i.e., the DNA can be the genetic material.

Though the above experiments provided strong evidence in favour of DNA as the genetic material, it was not clear DNA molecule is the repository of genetic information. Subsequent studies made by Erwin Chargaff, Maurice Wilkins, Rosalind Franklin, James Watson and Francis Crick led to the discovery of DNA structure, clarifying how DNA can encode large amounts of information (described in Chapter 3).

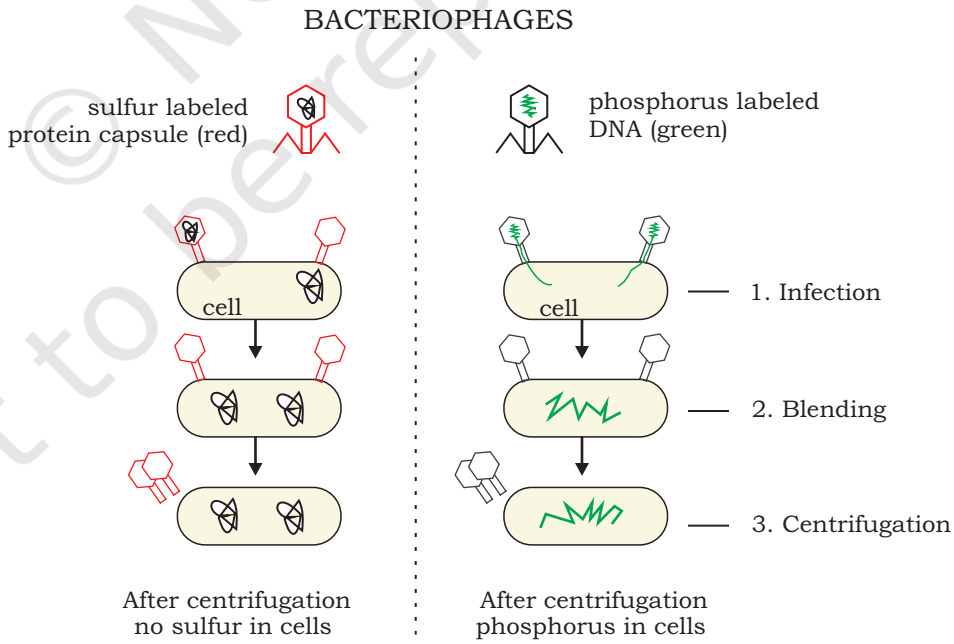


Fig. 7.3: Hershey-Chase experiment

© NCERT not to be republished

7.2 Prokaryotic and Eukaryotic Gene Organisation

It is well understood that traits get inherited from parents to offspring as 'gene unit' and DNA is the genetic material in all organisms except some viruses (where the genetic material is RNA). This led to the question about the organisation of gene, whether this organisation is similar in prokaryotes as well as eukaryotes and how does it function at the molecular level?

Gene Organization in Prokaryotes

There is no membrane bound nucleus in prokaryotes, its genetic material (i.e., circular double stranded DNA) is present in the specialised region within the cell called nucleoid region. Since the DNA molecule is very large and circular, question comes that how a large circular DNA is accommodated within the very small space of the cells? The organization and packaging of genetic material takes place through an important mechanism to accommodate the large DNA in small space of cell.

In prokaryotes, the genome is composed of a one, double-stranded DNA molecule in the form of a **loop** or **circle** (i.e. double stranded circular DNA). Some prokaryotes also have separate smaller loops of DNA (in addition to main nucleoid DNA) called **plasmids** that are not essential for normal growth. Thus, one-way prokaryotes compress their DNA into smaller spaces is through **supercoiling**. [Imagine twisting a rubber band so that it forms tiny coils. Now twist it even further, so that the original coils fold over one another and form a condensed ball]. When this type of twisting happens to a bacterial genome, it is known as **supercoiling**. Prokaryotic (bacterial) genomes can be **negatively supercoiled** (meaning that the DNA is twisted in the opposite direction of the double helix) or **positively supercoiled** (meaning that the DNA is twisted in the same direction as the double helix). **Most bacterial genomes are negatively supercoiled during normal growth.**

Multiple proteins act together to condense the prokaryotic DNA. In particular, **HLU** (histone like protein) is the most abundant protein in the nucleoid.

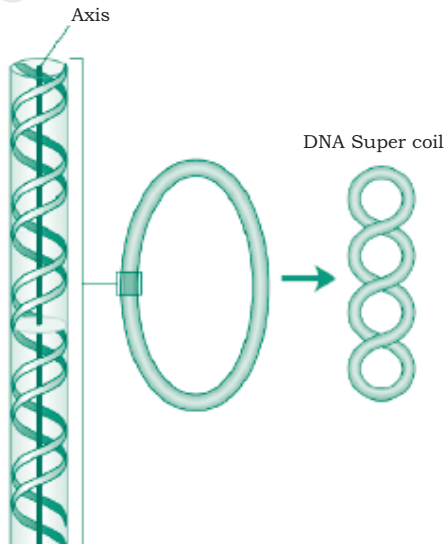


Fig. 7.4: Supercoiling of DNA in Prokaryote. When the axis of the double helix is coiled on itself, it forms a new helix (i.e. superhelix). The superhelix is usually called super coil.

© NCERT not to be republished

Gene Organization in Eukaryotes

Eukaryotes have well-defined membrane-bound nucleus within which genome is present in the form of **double stranded linear DNA molecules**. Since eukaryotic DNA is too long, it employs different packaging strategy to accommodate it inside the nucleus. At the most basic level, DNA is wrapped around the basic proteins known as **histones** to form the structures called **nucleosomes**. The histones are composed of **basic amino acids**; and its eight different molecules that associate to form an **octamer**. The DNA (which is acidic and negatively charged because of the presence of phosphate groups) is wrapped tightly around the basic core of histone octamer to form the spherical bead-like structure called **nucleosome**. Nucleosomes then appear as the beads on string (Fig. 7.5). One nucleosome is linked to the next one with the help of small segment of DNA called **linker DNA**.

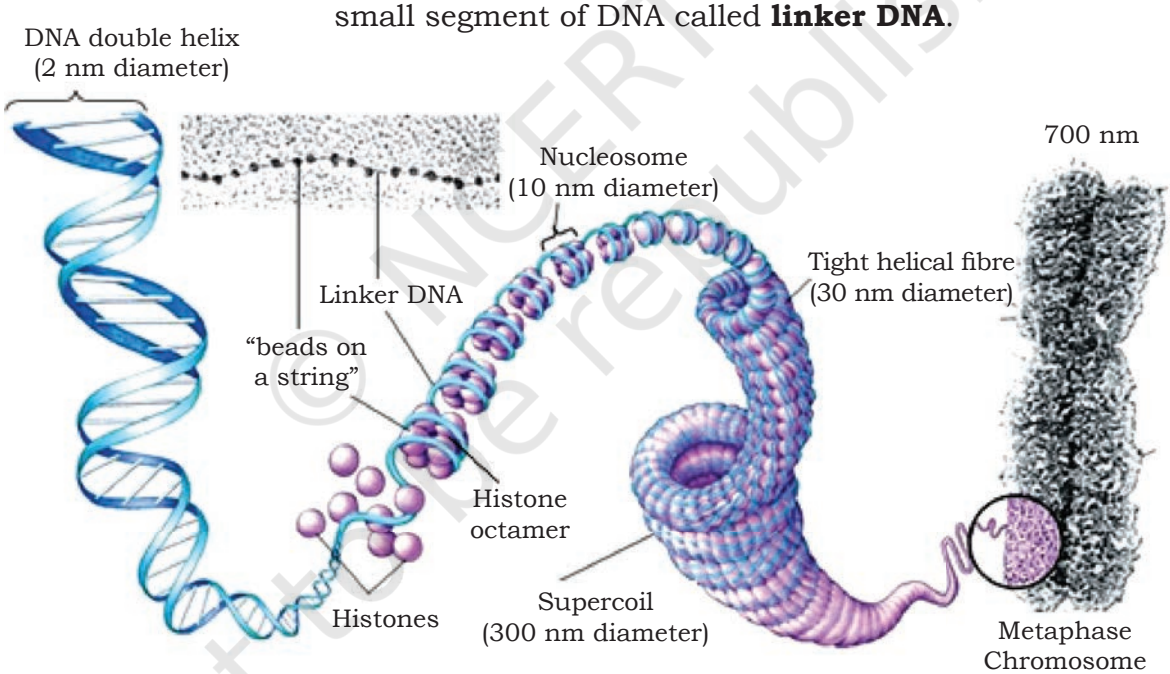


Fig. 7.5: Packaging of Eukaryotic Gene (Note: One chromosome is made up of only one DNA.)

The **core histone octamer** of each nucleosome consists of eight histone molecules, two each of four different histone types: H2A, H2B, H3, and H4. However, H1 is not a part of histone core assembly, but it binds independently with **linker DNA** and hence called as **linker**

© NCERT not to be republished

histone. Only single H1 histone is present on each linker DNA;. H1 histone is slightly larger than other histones.

Although nucleosomes may look like extended "beads on a string" under an electron microscope, they appear differently in living cells. In such cells, nucleosomes (10 nm diameter) stack up against one another in an organized manner with multiple levels of packaging, and compacted into a fibre (30 nm diameter). These 30 nm fibres then form a series of loops, which fold back on themselves for additional compacting with scaffold (300 nm diameter). At the metaphase stage, the chromosomes are condensed to the maximum, are approximately 700 nm in width, and are found in association with scaffold proteins.

The total DNA content in one haploid set of or chromosomes (as well as organelle DNA) is called a genome. As we see in the case of viruses or bacteria, the size of genome is comparatively much smaller than that of eukaryotic genome. Eukaryotic genomes are much more complex than prokaryotic genomes. Plant genomes are even more complex than any other eukaryotic genomes. Estimated size of genome in various groups of living organisms is shown in Fig. 7.7. In most of the eukaryotes a major part of the genome remains unexpressed.

In early 1940s, the cytological investigations of chromatin fibre revealed somewhat bead on a string like structure (Fig. 7.6) through electron microscopy and it was readily concluded that each bead perhaps represents a gene. Later investigations revealed that each bead is a nucleosome (containing a core of histone octamer and a double stranded DNA of 146 bp) and a string between the two beads, the linker DNA. It was also established that each nucleosome with its linker region involves approximately 200 bp. This cannot be considered to be a gene, as the size of the genes in many cases has to be much bigger than 200 nucleotides. The simple reason is that many proteins have more than 100 amino acid residues

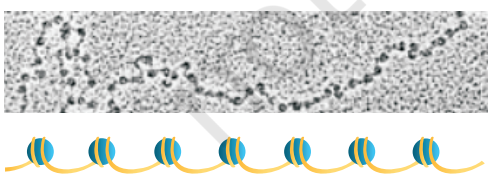


Fig. 7.6: Beads on string structure of chromatin

© NCERT not to be republished

and their corresponding regulatory genes cannot be less than three times of the same (based on triplet nature of the codon).

It has now become evident that a gene is the segment of DNA with specific promoter region, where RNA polymerase can bind and transcribe mRNA. The transcribed mRNA then gets involved in the process of translation. The mechanism is same for all organisms from virus to bacteria, plants and animals. Are genes of a virus, bacterium or higher organism similar in their structure and function?

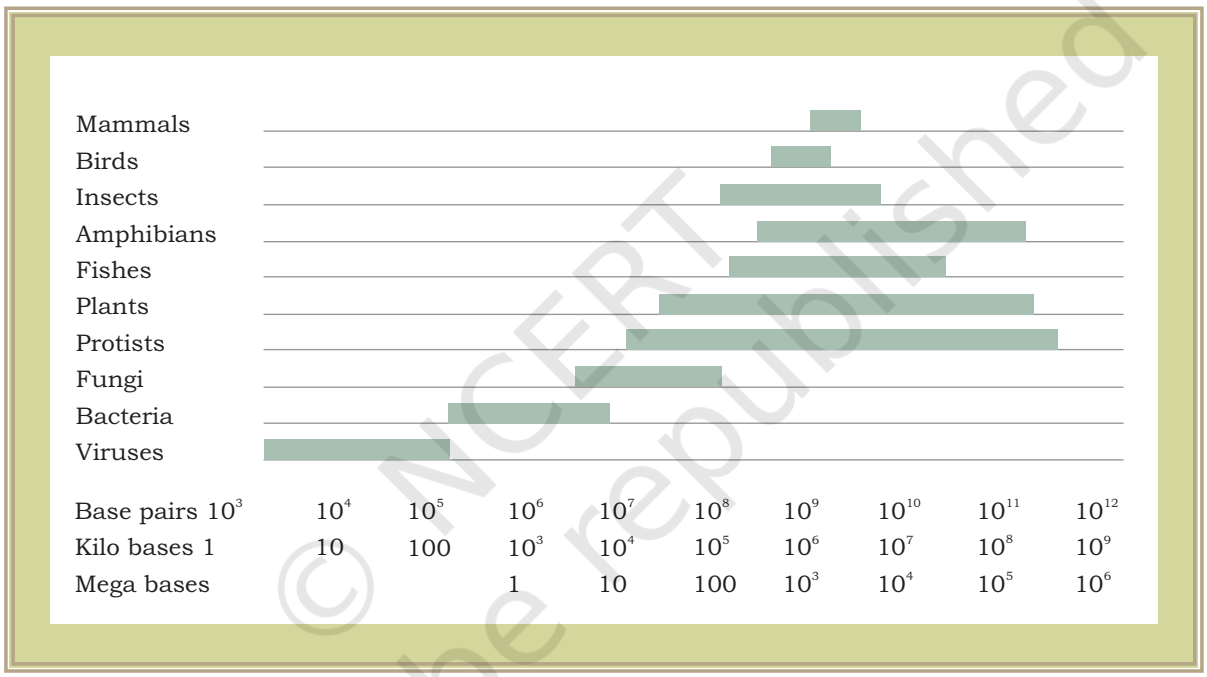


Fig. 7.7: Genome size variations in various groups of living organisms. Genome sizes are measured in thousands of nucleotide pairs, i.e., 1000 bp = 1 Kilobase (kb) and 1000,000 bp = 1000 Kb = 1 Megabase (Mb)

In most of the eukaryotes a major part of the genome is not expressed and remains as non-coding sequence. It has also been observed that in eukaryotic gene expression the size of active mRNA involved in the process of polypeptide chain synthesis is much smaller than the primary transcript. In fact, many of the eukaryotic genes e.g., the β-globin gene of haemoglobin after transcription undergoes the process of splicing in which a few interspersed segment

© NCERT not to be republished

of primary transcript is excised (introns) and remaining portions (exons) of the RNA transcript are joined together to form mRNA.

A eukaryotic cell has two types of genome: (i) nuclear genome and (ii) organellar genome.

Nuclear Genome

In eukayotes, Majority of the DNA is found in the nucleus and is known as nuclear DNA. In prokaryotes, most of the genome is composed of coding DNA sequences while in eukaryotic genomes coding regions makes relatively very small part of the total genome. For example, size of the human genome is about 3,000 Mb or 3 billion base pairs of DNA and estimated to have more than 20,000 genes, which constitutes approximately 2% of the total genome. In non-coding region of the genome, there are sequences which are repeated thousands to several million times in the form of tandem arrays. Size and number of these repetitive DNA sequences in the genome varies significantly.

Organellar Genomes

In addition to nuclear DNA, few membrane-bound cellular organelles like chloroplast and mitochondria contain organelle-specific DNA. Organelle genomes are mostly circular double stranded DNAs and are present in multiple copies in each organelle (Fig. 7.8). These replicate in semi-conservative fashion and are inherited separately from the nuclear genome. Organelle DNA contains genes that are required for organelle-specific functions and are usually uniparental and inherited to next generation through female gametes.

7.3 DNA Replication

When the three dimensional structure of DNA was proposed by James Watson and Francis Crick in 1953, the feature that most excited the biologists was the complementary relationship of bases of two polynucleotide chains. Watson and Crick immediately suggested a model that the basis for copying the genetic information is complementarity (For details refer to Unit II, Chapter 2). According to the

© NCERT not to be republished

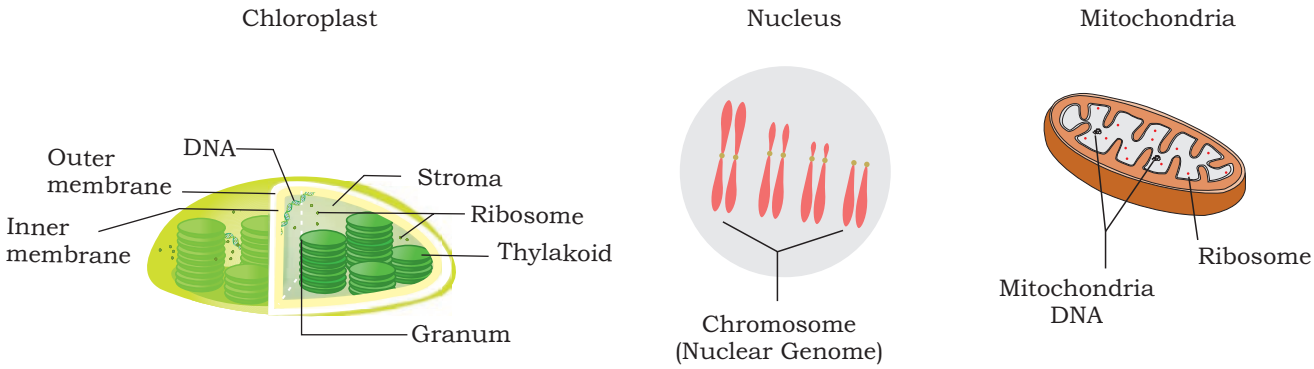


Fig. 7.8: Organellar DNA

model proposed by them, the two strands of DNA separate during replication; each strand serving as a template for the synthesis of a new complementary strand because of the specificity of base pairing (i.e., Thymine with Adenine and Cytosine with Guanine). Thus the parental duplex DNA to form two identical daughter duplex, each of which consists of one parental strand and one newly synthesised daughter strand. This form of DNA replication is called **semiconservative replication** (Fig. 7.9). The evidence for this mode of replication was provided by Mathew Messelson and Franklin Stahl in 1958.

7.3.1 Messelson and Stahl experiment

To distinguish between old and new strand, Messelson and

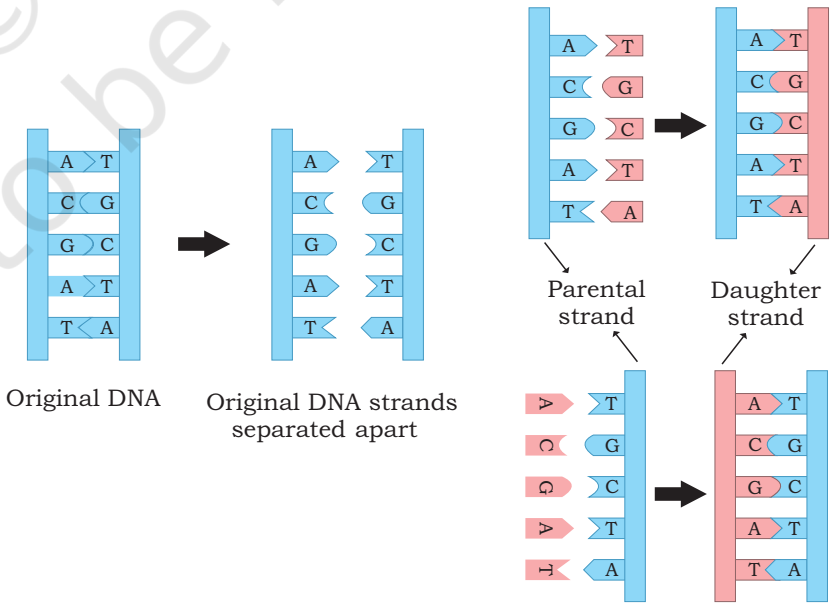


Fig. 7.9: Semiconservative mode of DNA replication

© NCERT not to be republished

Stahl used two isotopes of nitrogen, ^{14}N (the common form) and ^{15}N (a rare, heavy form). They grew *E. coli* bacteria in a medium containing the heavier isotope of nitrogen as the sole nitrogen source. After several generations, all the *E. coli* cells had ^{15}N incorporated into the purine and pyrimidine bases of DNA. By using density gradient centrifugation, they observed that the DNA extracted from *E. coli* grown in ^{15}N produced a single band at the lower side of the centrifuge tube, while DNA extracted from bacterial cells grown in ^{14}N medium formed a band closer to the top. This indicated that the DNA of *E. coli* cells grown in ^{15}N medium was denser than that of bacteria grown in a medium containing the lighter isotope (^{14}N) (Fig. 7.10).

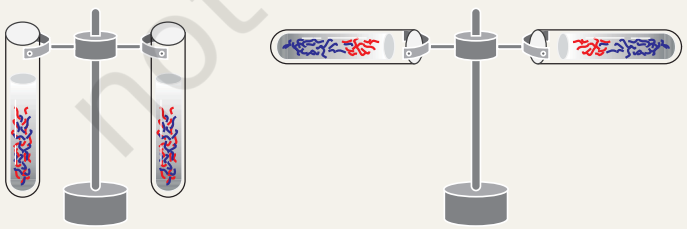
Messelson and Stahl then transferred the *E. coli* bacteria from the ^{15}N medium to the ^{14}N medium and collected DNA at various time intervals as the bacterial cells multiplied (*E. coli* divides after every 20 minutes). The DNA extracted from *E. coli* cells after first round of division (generation I) when analysed by using cesium chloride salt in density gradient

Mathew Messelson and Franklin Stahl confirmed semi-conservative mode of DNA replication

Box 1

Density Gradient Centrifugation

It is a centrifugation technique for separating the molecules from a mixture on the basis of their density. The centrifugation tube is filled with a heavy salt solution of cesium chloride (CsCl) and the sample whose density is to be measured. The centrifuge tube is then allowed to spin in an ultracentrifuge at a very high speed for several days. The enormous artificial force generated by the ultracentrifuge causes the Cs ions to migrate towards the bottom of the tube, creating a gradient with high density at the bottom and low density at the top. The DNA strands float or sink in the gradient till the density matches the density of the salt.



It is then spun in a centrifuge at high speed for several days

A centrifuge tube is filled with a heavy salt solution and DNA fragments

A density gradient develops within the tube. Heavy DNA (with ^{15}N) will move towards the bottom; light DNA (with ^{14}N) will remain closer to the top

© NCERT not to be republished

centrifugation) produced a single band, but at a position intermediate between the bands of heavy DNA (^{15}N) and light DNA (^{14}N) bands. When DNA was extracted from *E. coli* cells, after a second round of replication in ^{14}N medium (generation II) two bands of equal intensity appeared in the centrifuge tube, one in the intermediate position and other at the position expected of DNA that contained only ^{14}N DNA. When samples of DNA were collected after additional rounds of replication they produced two bands. The bands representing light DNA became progressively thicker but the band at the intermediate position remained unchanged.

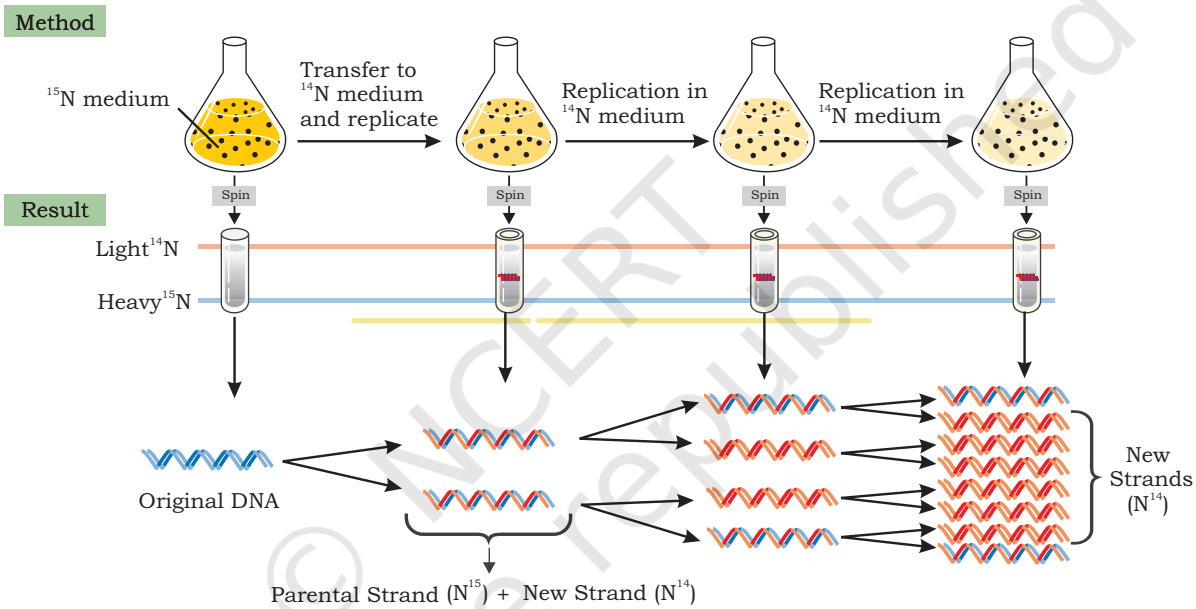


Fig. 7.10: Messelson - Stahl experiment to confirm semiconservative mode of DNA replication

7.3.2 Interpretation of results by Messelson and Stahl

After the first round of replication, each daughter DNA duplex was a hybrid having one heavy strand containing ^{15}N from the parent and one light strand containing ^{14}N from the medium. When this hybrid duplex replicated, the heavy template strand formed another hybrid duplex DNA while the lighter formed light DNA duplex. Thus, Messelson and Stahl experiment clearly confirmed the prediction of Watson and Crick model that DNA replicates in a semi-conservative manner.

Taylor and colleagues in 1958 also conducted similar experiments in *Vicia faba* by using radioactive thymidine

© NCERT not to be republished

to detect the distribution of newly synthesised DNA in the chromosomes. They also proved that DNA in chromosomes replicate semiconservatively.

7.3.3 The machinery of replication

In order to carry out the replication of double stranded DNA molecule it must unwind for separating the strands to expose the bases. Both DNA strands act as template for the assembly of complementary bases to form new polynucleotide strands which are anti-parallel to the template strand. Several proteins and enzymes are required for unwinding the double helix and synthesis of new strands of DNA. The key enzymes and proteins involved in the replication of DNA in both prokaryotes and eukaryotes include—

(i) *DNA Polymerases*

These are the main enzymes of replication as they are responsible for synthesising new DNA strands. They synthesise new DNA strands in 5'→3' direction by catalysing the formation of phosphodiester bond between the 3′ hydroxyl group at the growing end of DNA chain and the 5′ phosphate group of incoming deoxy-ribonucleoside triphosphate (dNTP).

In prokaryotes, there are three kinds of DNA polymerases—DNA polymerase I, II, III. DNA polymerase III is the main enzyme of DNA replication. It synthesises new strand in 5'→3' direction and by its 3'→5' exonuclease activity that enables it to correct errors in the growing chain by removing mis-incorporated nucleotides in 3'→5' direction. The DNA polymerase I has 5'→3' polymerisation activity in addition to both 5'→3' and 3'→5' exonuclease activity. By its 5'→3' exonuclease activity, it removes RNA primers laid down by primase. The DNA polymerase II is involved in repair of DNA. In eukaryotes, multiple DNA polymerases are involved in synthesis of new strands and repair of damaged DNA.

(ii) *Primase*

The drawback of DNA polymerase is that it cannot initiate synthesis of a new strand of DNA. It needs a RNA primer, an existing segment of nucleotides that provides 3′ OH group to which it can add a new nucleotide. The primase,

© NCERT not to be republished

a DNA dependent RNA polymerase synthesises short RNA oligonucleotide chain (about 10-12 nucleotides long) called RNA primer in order to get DNA replication started.

(iii) *Helicases*

In order to generate single strand templates, the DNA double helix must unwind. DNA helicase breaks hydrogen bonds that exist between the bases of two strands of a DNA molecule by using energy in the form of ATP.

(iv) *Topoisomerase*

As the two strands of DNA separate, torsional strain causes the DNA helix to coil up forming a knot in front of the replication fork. Topoisomerases relieve the tension by nicking and religating the DNA that holds the helix in its coiled and supercoiled structure.

(v) *Single strand binding (SSB) proteins*

After unwinding of DNA by helicase, the single stranded nucleotide chains have a tendency to form hydrogen bonds again and reanneal. SSB proteins attach tightly to the exposed single stranded DNA and stabilise them in single stranded form for replication to take place.

(vi) *DNA ligase*

DNA ligase enzyme joins the newly synthesised DNA fragments by forming phosphodiester bond between 3′OH group and 5′ phosphate group.

7.3.4 Mechanism of DNA replication

In prokaryotes, replication usually starts at a specific site on the DNA sequence known as **origin of replication (*ori*)** (Fig. 7.11). In bacteria, DNA is circular and double stranded. It has single replication origin (in *E. coli* called *oriC*) and ends at a specific site, knows as terminus. Initiator proteins bind to origin of replication and unwind a short section of DNA. Further, separation of two strands on both sides of initial opening is brought about by helicase, by breaking the hydrogen

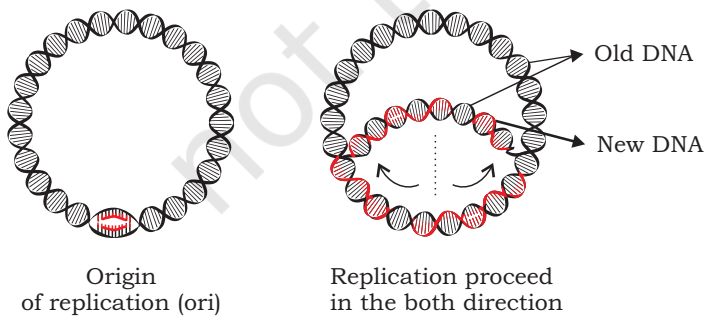


Fig. 7.11: Bacterial chromosomes have a single point of origin

© NCERT not to be republished

bonds present between the bases of two nucleotide strands. Replication forks are formed on both sides and they move away from the origin in opposite direction. This mode of replication is called bidirectional replication (Fig. 7.11). As DNA synthesis requires single stranded template, single strand binding proteins stabilise the single stranded DNA by binding to it. A topoisomerase, reduces torsional strain generated in front of replication fork as a result of unwinding of DNA strands (Fig. 7.12).

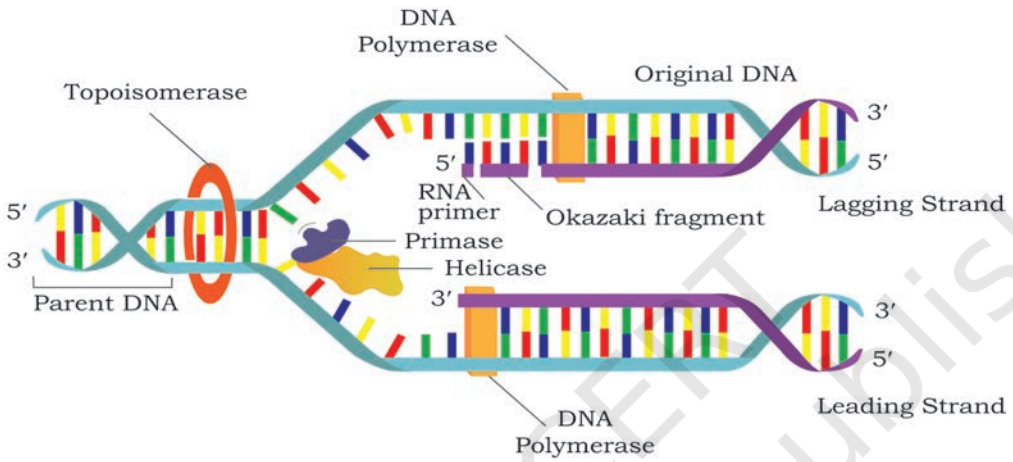


Fig. 7.12: DNA synthesis takes place simultaneously but in opposite directions on the two DNA template strands

DNA polymerase III synthesises a new DNA strand in 5'→3' direction, antiparallel to the template strand (Fig. 7.13). DNA polymerase cannot initiate DNA synthesis on a bare template; rather can add nucleotide to 3'OH group of a primer strand. The primase synthesises a RNA primer, a short stretch of about 10-12 nucleotides long in 5'→3' direction on template strand in each replication fork at the origin. The RNA primer provides 3'OH group to which DNA polymerase add nucleotides. After the formation of RNA primer on the template strand oriented in 5'→3' direction in the replication fork, DNA polymerase III elongates the polynucleotide strand by catalysing the formation of phosphodiester bond between the 3'OH group present at the growing end of DNA strand and 5' α phosphate group of incoming deoxyribonucleoside triphosphate (dNTP). In each step, β and γ phosphates of incoming dNTP are cleaved and the resulting nucleotide (nucleotide monophosphate) is added to the 3'-OH group

© NCERT not to be republished

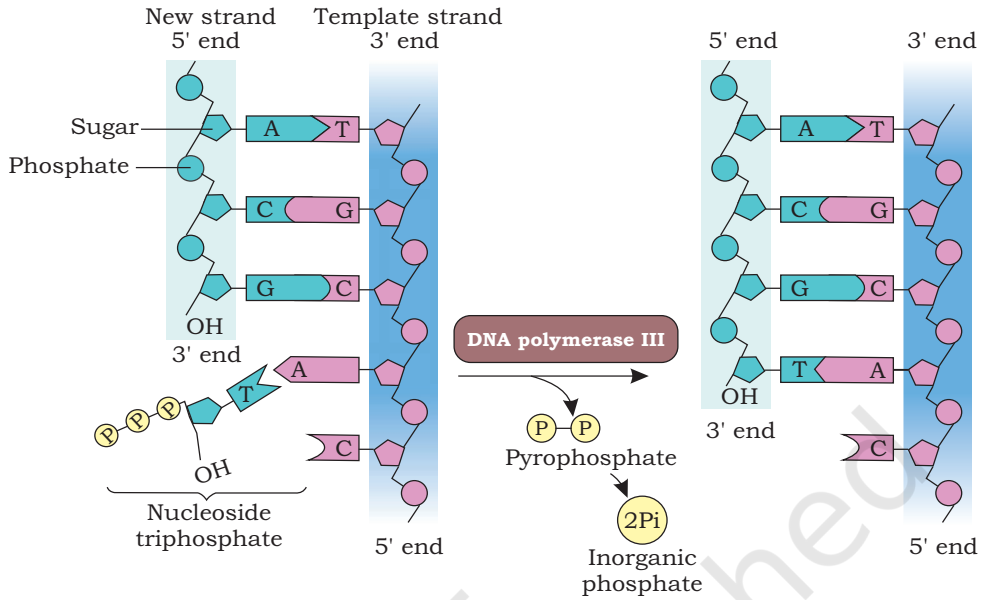


Fig. 7.13: DNA polymerase catalyses the addition of the 5' phosphate group to the 3' OH group of the previous nucleotide

of the growing nucleotide strand. On the template strand, oriented in 3'→5' direction, the new strand is synthesised continuously in 5'→3' direction and is called as leading strand (Fig. 7.14).

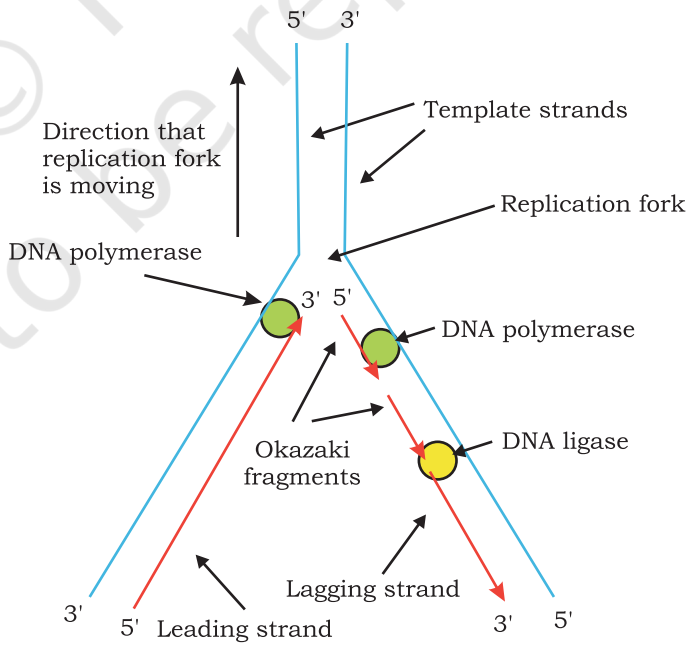


Fig. 7.14: Leading strand is synthesised continuously and lagging strand synthesised discontinuously

© NCERT not to be republished

Synthesis of new DNA strands on both template strands in replication fork takes place simultaneously. On the template strand, oriented in 5'→3' direction in replication fork, replication proceeds in opposite direction to that of the movement of replication fork. The primase synthesises a RNA primer at the replication fork on this template strand. The DNA polymerase III adds nucleotide to the 3′OH end of primer and synthesises short stretch of DNA of about 1000 – 2000 nucleotides. As unwinding proceeds, another RNA primer is formed which is elongated by DNA polymerase. Synthesis on this template takes place discontinuously in the form of short segments called Okazaki fragment named after R. Okazaki, who along with colleagues first identified them. Each Okazaki fragment begins with an RNA primer. The DNA strands synthesised discontinuously is called lagging strand.

Since during DNA replication, one strand is synthesised continuously (leading strand) and the other strand is synthesised discontinuously (lagging strand), replication is said to be semi-discontinuous. Thus, leading strand has only one RNA primer at its 5' end while lagging strand has multiple RNA primers (equal to number of Okazaki fragments). The question here arises is, how are these Okazaki fragments converted into a continuous DNA strand? DNA polymerase I by its 5'→3' exonuclease activity removes nucleotides of RNA primer and replaces them with complementary DNA nucleotides by its 5'→3' polymerisation activity. The DNA ligase joins the Okazaki fragments by catalysing formation of phosphodiester bonds between them.

The exact details of the termination process are not clear, but it is known that a specific termination site is located roughly opposite to origin of replication on the circular DNA molecule. Replication is terminated whenever two replication forks meet (Fig. 7.15). A termination protein binds to termination site and blocks the movement of helicase thus stopping the replication fork and preventing further DNA replication.

The eukaryotic DNA replication resembles bacterial replication in many respects. Eukaryotes have multiple chromosomes containing linear, double

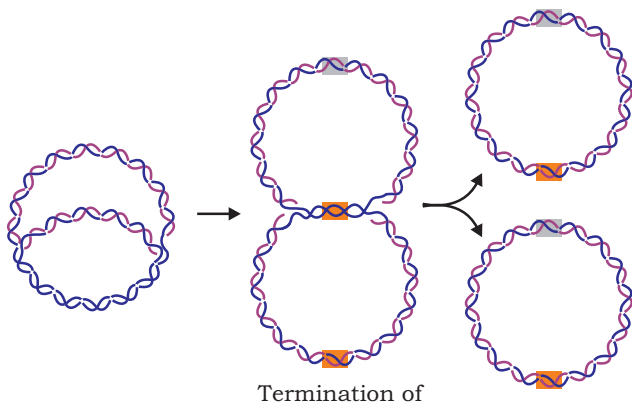


Fig. 7.15: Termination of replication

© NCERT not to be republished

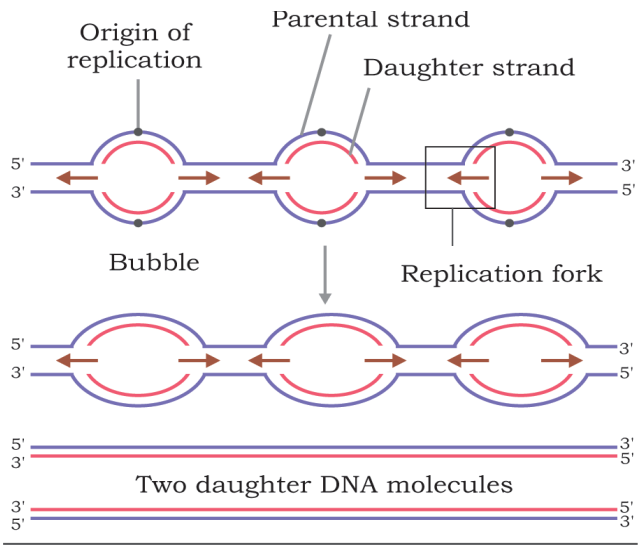


Fig. 7.16: Eukaryotic chromosomes have multiple points of origin

stranded, long DNA molecules. Instead of single origin, each eukaryotic chromosome has multiple origins and replication forks are bidirectional. Replication can start simultaneously from all origins. Several kinds of DNA polymerases are found in eukaryotes involved in replication and repair (Fig. 7.16).

7.4 GENE EXPRESSION

As we have already studied that gene is a segment of DNA which carries biological information for the expression of a trait. One of the great challenges in understanding gene is how the gene is expressed. In other words, how is this information in the form of linear sequence of nucleotides in a polynucleotide chain converted into the linear sequence of amino acids in a polypeptide chain?

Every organism from bacteria to human beings has the same basic mechanism of expression of genes (Fig. 7.17). The information encoded in the sequence of the four bases of DNA directs the assembly of amino acids in the correct order, so as to produce the protein for which the given DNA sequence is responsible. The DNA inherited by an organism leads to specific traits by directing the synthesis of specific proteins. In other words, proteins are the link between genotype and phenotype. This unidirectional flow of information from DNA to proteins involves two steps, i.e., transcription and translation and is called **central dogma**.

During transcription the genetic information is transferred from DNA to mRNA. In the second step of central dogma, i.e., translation the information is transferred from mRNA to polypeptide chain. Why would the cell want to have mRNA as an intermediate between a gene present in DNA and the peptide it encodes? First of all, the information present in gene can be amplified by having many copies of mRNA made from a single copy of DNA. Second, in eukaryotes, DNA is present inside nucleus

© NCERT not to be republished

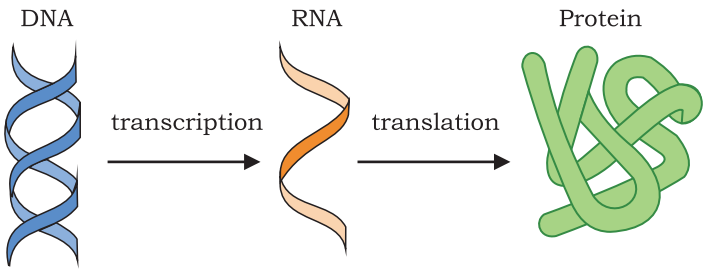


Fig. 7.17: Central dogma – unidirectional flow of information

$$\text{DNA} \xrightarrow{\text{Transcription}} \text{RNA} \xrightarrow{\text{Translation}} \text{Polypeptide chain}$$

and ribosomes, the protein factories are present in cytoplasm. Once the mRNA is synthesised as a complementary copy of DNA strand (gene), it moves into cytoplasm. In the cytoplasm, it serves as a template for the synthesis of polypeptide chain in ribosome (Fig. 7.18).

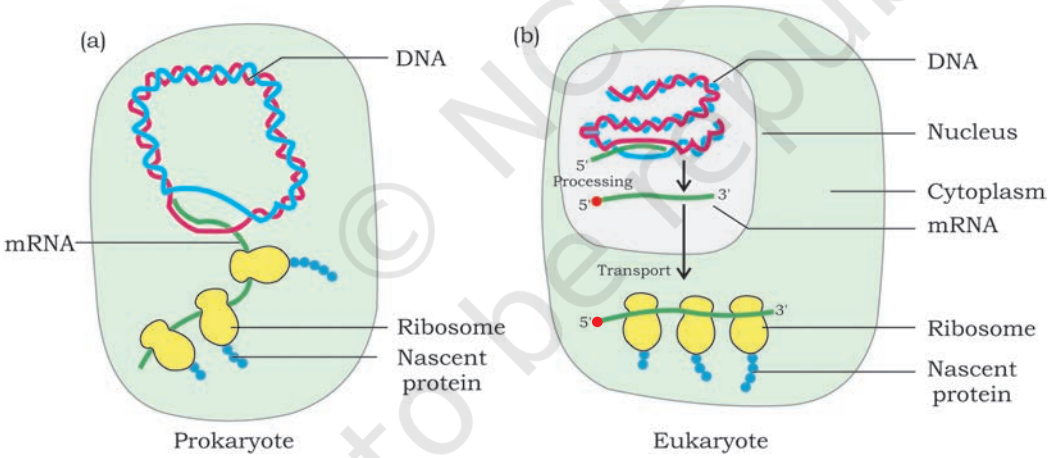


Fig. 7.18: Gene expression in prokaryotes and eukaryotes

In some viruses like retroviruses, the genetic material is RNA instead of DNA. The information present in genetic RNA is transferred to a single stranded complementary DNA strand which is then converted into double stranded DNA. This process is called **reverse transcription**.

© NCERT not to be republished

After the formation of DNA, the genetic information flows to mRNA and then to polypeptide chain (Fig 7.19).

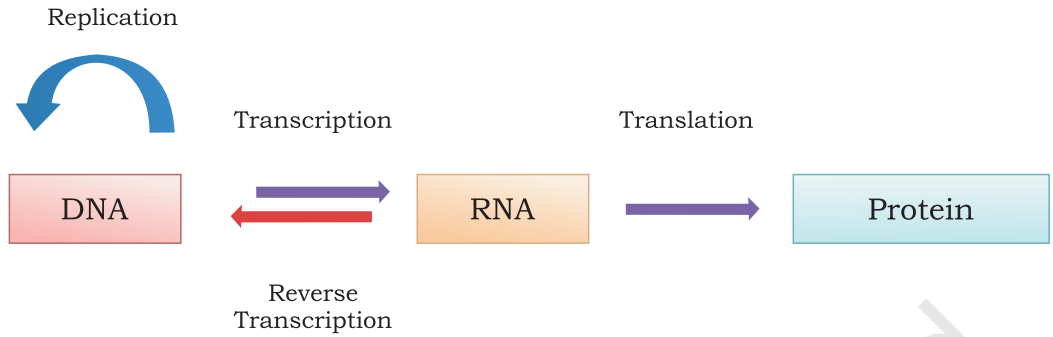


Fig. 7.19: Reversal of central dogma of molecular biology

7.4.1 Transcription

All kinds of cellular RNAs such as messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA) are synthesised on DNA template. Synthesis of complementary RNA polynucleotide chain on DNA is called transcription. The transcription is in many ways similar to the process of replication. One fundamental difference is that during replication, the entire DNA template is copied to DNA molecule but, during transcription, only small parts of the DNA are copied into RNA. The DNA strand whose base sequence is identical to that of the transcribed RNA (except for T in DNA to U in RNA) that carries genetic information is called sense or coding strand. The other DNA strand on which RNA is transcribed and whose nucleotide sequence is complementary to that of the transcribed RNA is called the template or non-coding strand (Fig. 7.20). Thus, within a gene, only one of the nucleotide strands is normally transcribed into RNA.

RNA polymerase

The enzyme responsible for transcription in both prokaryotes and eukaryotes is DNA-dependent RNA polymerase. In prokaryotes there is a single type of RNA polymerase that catalyses the transcription of all types of RNA (mRNA, rRNA and tRNA). In eukaryotes, three distinct types of RNA polymerases are present. RNA polymerase I transcribes 28S, 18S and 5.8S rRNA; RNA polymerase II transcribes hnRNA (heterogenous nuclear RNA, the precursor of mRNA) and

© NCERT not to be republished

RNA polymerase III transcribes tRNA and 5S rRNA. RNA polymerase enzyme do not require a primer for initiating synthesis of polynucleotide chain and to synthesise new chain in 5'→3' direction using ribonucleoside triphosphates (rNTPs) as substrates.

The prokaryotic RNA polymerase is a large enzyme comprising of two α, one β and one β' subunits (Fig. 7.20). These subunits together constitute **core enzyme** (α_2, β, β'). When core enzyme associates with a sigma factor it forms **holoenzyme** (α_2, β, β', σ).

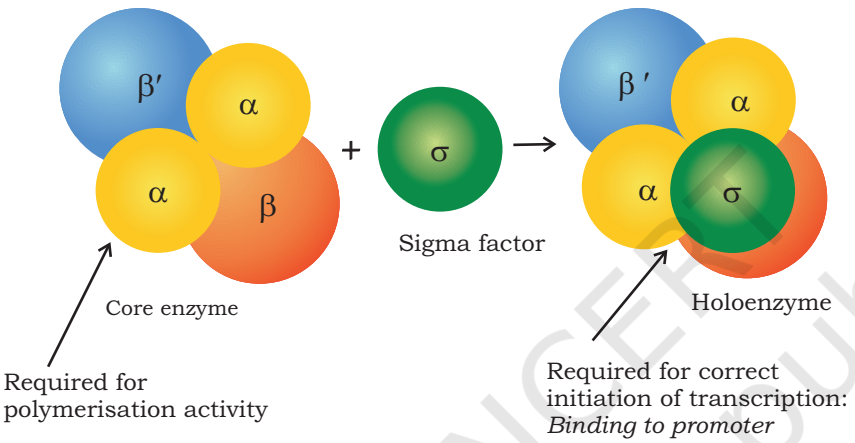


Fig. 7.20: Prokaryotic RNA polymerase

Transcription unit

The stretch of DNA on which RNA is transcribed is called a transcription unit. How does the RNA polymerase recognise a transcription unit? How does it know which DNA strand to transcribe and where to start and stop? Each transcription unit has a start site from where transcription start and a terminator site where transcription is to end (Fig. 7.21). Upstream to the start site is a DNA sequence called promoter, which is recognised by the RNA polymerase and binds to it for accomplishing initiation of transcription. In addition to providing the binding site for the polymerase, the promoter also tells the polymerase where to start synthesis and in which direction to proceed. Within promoter is a most common consensus sequence TATAAT called **Pribnow box** in bacteria where initial melting of duplex DNA takes place.

© NCERT not to be republished

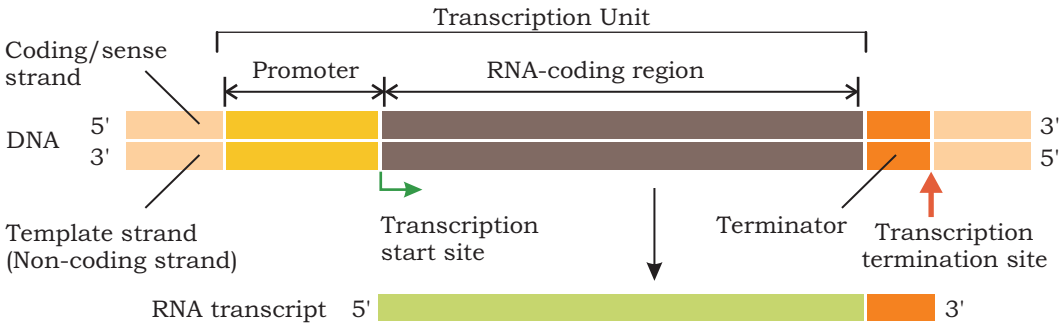


Fig. 7.21: A transcription unit

Initiation

The binding of RNA polymerase to the promoter is the first step in transcription. In bacteria, sigma subunit of RNA polymerase recognises promoter and binds to it. Once bound to promoter, the RNA polymerase holoenzyme begins to unwind the DNA helix at the **TATAAT sequence**. It unwinds a DNA segment approximately 17 base pairs long. To begin the synthesis of an RNA molecule, RNA polymerase pairs the base of an incoming ribonucleoside triphosphate with complementary base at the start site on the DNA template strand. No primer is required to initiate the synthesis of the 5' end of the RNA molecule. The next ribonucleoside triphosphate complementary to the second nucleotide of template strand is added to the 3′ OH end of the first RNA nucleotide by a phosphodiester bond catalysed by RNA polymerase. During this, the two (β and γ) of the three phosphate groups are cleaved as pyrophosphate from the incoming ribonucleoside triphosphate. The sigma subunit is usually released after initiation.

Elongation

The region containing the RNA polymerase, DNA and growing RNA transcript is called transcription bubble because it contains a locally unwound 'bubble' of DNA (Fig. 7.22). The RNA polymerase moves along the template strand progressively and unwind the DNA at the leading edge of the transcription bubble. It joins nucleotides to the 3'OH end of RNA molecule complimentary to the sequence of the template (Fig. 7.23). As the elongation of RNA chain continues, the new RNA molecule peels away from the DNA template and the DNA double helix reforms behind

© NCERT not to be republished

transcription bubble. The rate of transcription in bacterial cells at 37^{0}C is about 40 nucleotides per second.

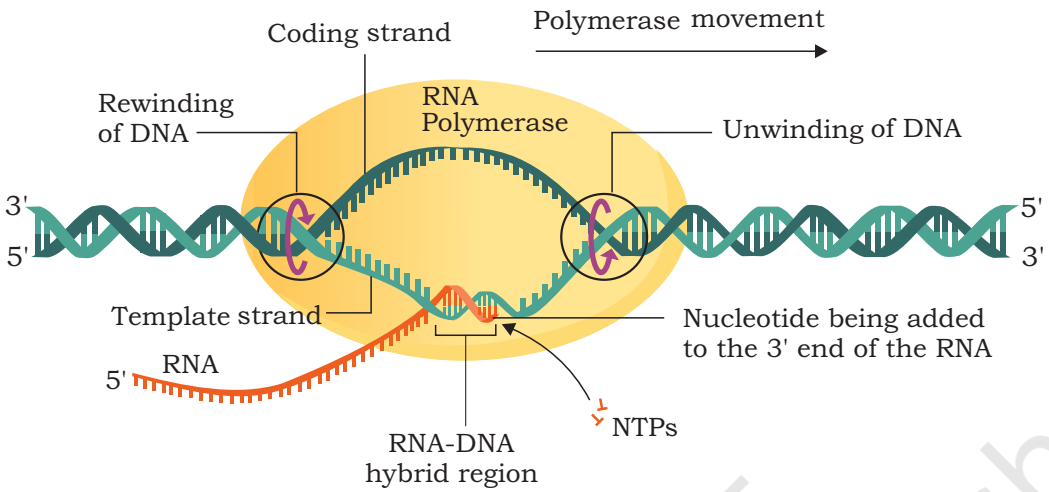


Fig. 7.22: Elongation of RNA chain by RNA polymerase

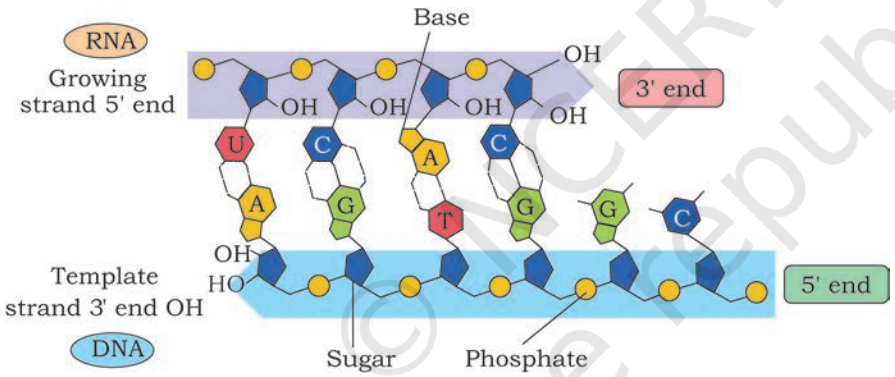


Fig. 7.23: Addition of NTP to the 3'OH end of the growing polynucleotide chain

Termination

The elongation of RNA chain continues till the RNA polymerase reaches the terminator sequence in DNA. The RNA-DNA hybrid helix within transcription bubble dissociates as RNA polymerase reaches terminator sequence. The RNA polymerase separates from template DNA, the two strands of DNA rewind and newly synthesised RNA chain gets released. In prokaryotic termination of transcription, RNA polymerase requires a rho protein (rho dependent termination) in some cases, while in others it is terminated with rho independent termination.

© NCERT not to be republished

In prokaryote, a group of genes is often transcribed into a single RNA molecule, called polycistronic RNA. It is produced when a single termination sequence is present at the end of a group of several genes that are transcribed together. In prokaryote, since the mRNA does not require any processing to become active, and also since transcription and translation take place in the cytoplasm, many times the translation can begin much before the mRNA is fully transcribed.

Transcription in eukaryotes

The basic mechanism of transcription by RNA polymerase is the same in eukaryotes as in prokaryotes. However, a number of differences in the process of transcription occur between prokaryotes and eukaryotes. A number of accessory factors known as trascription factors are required for proper initiation of transcription by RNA polymerase I, II and III at promoters of rRNA, mRNA and tRNA genes respectively. Most of the genes in eukaryotes

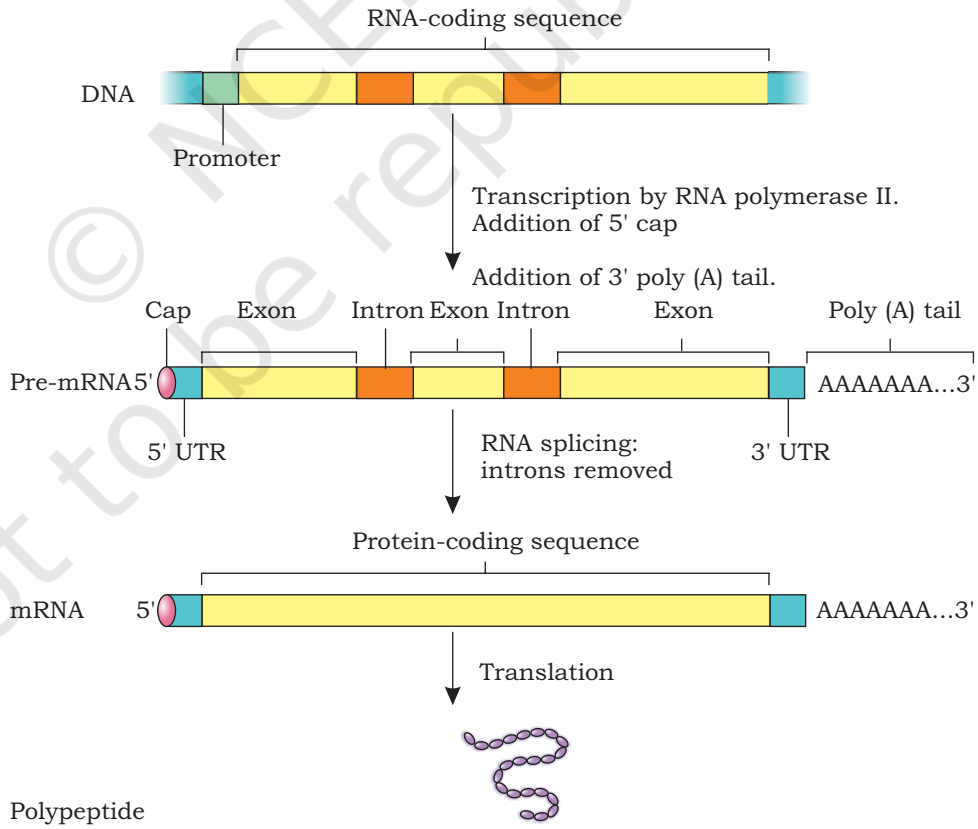


Fig. 7.24: Post-transcriptional modifications of pre-mRNA

© NCERT not to be republished

are split genes containing introns (non-coding regions) in between exons (coding regions). The primary transcripts contain both the exons and the introns. The primary transcripts called heterogenous nuclear RNA (hnRNA) undergo processing before getting transferred into cytoplasm from nucleus called **post-transcriptional modifications** (Fig. 7.24).

1. Capping

In eukaryotic pre-mRNA, the first nucleotide is a purine (A or G) at the 5' end. The 5' end of the pre-mRNA is modified by the addition of GTP called a 5' cap. The guanine in the GTP is also modified by the addition of a methyl group often called 5' methyl G cap. The cap protects 5' end of mRNA from degradation by exonuclease.

2. Splicing

The primary transcripts or pre-mRNA has both exons and introns. During processing of pre-mRNA, the introns are cleaved and exons are joined (spliced) together. This process is called splicing.

3. Poly-A tail

After the transcription of pre-mRNA, a series of adenine nucleotides are added to the 3' end called poly-A tail or poly-adenylated tail. The poly-A tail appears to play a role in the stability of mRNA by protecting them from degradation.

7.5 Genetic Code

In case of replication and transcription, a polynucleotide chain (template strand) is copied to form another, polynucleotide strand, i.e., a DNA strand or an RNA strand respectively. These processes are easy to conceptualise on the basis of complementarity. But in the process of translation, genetic information is transferred from a polymer of nucleotides (mRNA) to a polymer of amino acids. No complementarity exists between nucleotides and amino acids. The question arises how the sequence of four bases in mRNA specifies the amino acid sequence of a polypeptide? Evidences suggest that a minute alteration in the nucleotide sequence is accompanied by a change in the amino acid sequences of the polypeptides. This led to the proposition of a genetic code that could direct the

© NCERT not to be republished

amino acid sequence during protein synthesis.

The proposition and deciphering of the genetic code was indeed challenging as it necessitated the collective involvement and efforts of physicists, organic chemists, biochemists and geneticists. In 1961, Francis Crick hypothesized the existence of a genetic code and suggested base sequence as the carrier of genetic information. How many nucleotides are necessary to specify a single amino acid? Since there are twenty different amino acids and only four different bases, it would be logically impossible for each amino acid to be specified by only one nucleotide. Similarly, combination of two nucleotides could also specify sixteen amino acids as there will be 16 codons. George Gamow, a physicist, argued that since there are 4 bases and if they have to code for 20 amino acids, the code should constitute a combination of bases. Further he added that in order to produce 20 amino acids there should be 3 nucleotides constituting a genetic code. This was indeed a bald proposition, because a permutation combination of 4^3 gives 4×4×4 that would generate 64 codons, which is more than enough to specify 20 different amino acids.

The next major step was to determine which groups of three nucleotides specify which amino acids. In 1961, the experiment conducted by Marshall Nirenberg and Johann H. Matthaei deciphered the first codon. They synthesised an artificial mRNA by linking RNA nucleotides containing uracil as their base. The poly (U) RNA was added to a test tube containing all twenty types of amino acids, ribosomes and the other components required for protein synthesis. In each test tube, a particular amino acid was made radioactive. A radioactive polypeptide chain was detected in one of the test tubes containing phenylalanine amino acids. Thus, Nirenberg and Matthaei determined that the mRNA codon UUU specifies the amino acid phenylalanine. The results of similar experiments using poly (C) and poly (A) RNA demonstrated that CCC codes for proline and AAA codes for lysine. The experiments of Nirenberg, Matthaei, Leder, Ochoa and H. G. Khorana helped in deciphering all 64 codons of genetic code.

The salient features of genetic code are as follows:

1. The codons are triplet and there are 64 codons (Fig. 7.25).

© NCERT not to be republished

First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU Phe, UUC Phe	UCU Ser, UCC Ser, UCA Ser, UCG Ser	UAU Tyr, UAC Tyr	UGU Cys, UGC Cys	U, C
	UUA Leu, UUG Leu		UAA Stop, UAG Stop	UGA Stop, UGG Trp	A, G
C	CUU Leu, CUC Leu, CUA Leu, CUG Leu	CCU Pro, CCC Pro, CCA Pro, CCG Pro	CAU His, CAC His	CGU Arg, CGC Arg, CGA Arg, CGG Arg	U, C
			CAA Gln, CAG Gln		A, G
A	AUU Ile, AUC Ile, AUA Ile	ACU Thr, ACC Thr, ACA Thr, ACG Thr	AAU Asn, AAC Asn	AGU Ser, AGC Ser	U, C
	AUG Met		AAA Lys, AAG Lys	AGA Arg, AGG Arg	A, G
G	GUU Val, GUC Val, GUA Val, GUG Val	GCU Ala, GCC Ala, GCA Ala, GCG Ala	GAU Asp, GAC Asp	GGU Gly, GGC Gly, GGA Gly, GGG Gly	U, C, A, G
			GAA Glu, GAG Glu		

Fig. 7.25: Genetic code

2. Out of 64 codons, 61 codons code for 20 amino acids. The remaining three codons UAA, UAG and UGA do not code for any amino acids and are used to signal the termination of protein synthesis.
3. AUG has dual functions. It codes for methionine and also acts as an initiator codon.
4. The genetic code is unambiguous, i.e., one codon codes for only one amino acid.
5. The genetic code is degenerate, which means that each amino acid may be specified by more than one codon. Only methionine and tryptophan are encoded by a single codon each.
6. The genetic code is non-overlapping. Each base along the mRNA is a part of only one codon.
7. The genetic code is universal; for example, from bacteria to human GUG would code for valine. Some exceptions to this rule have been found in mitochondrial codons, and in some protozoans.

7.6 Translation

So far you have learnt that genetic information present in DNA is transcribed into mRNA. Thus, mRNA carries genetic information for the sequence of amino acids

© NCERT not to be republished

of specific polypeptide chains in the form of codons. In ribosomes, information present in the language of codons is decoded into language of amino acids which join to form a polypeptide chain. The process of synthesis of polypeptide chain on an mRNA bound to ribosome is called translation.

The translation process can be divided into four steps:

(i) the charging of tRNA
(ii) the initiation of translation
(iii) elongation
(iv) termination

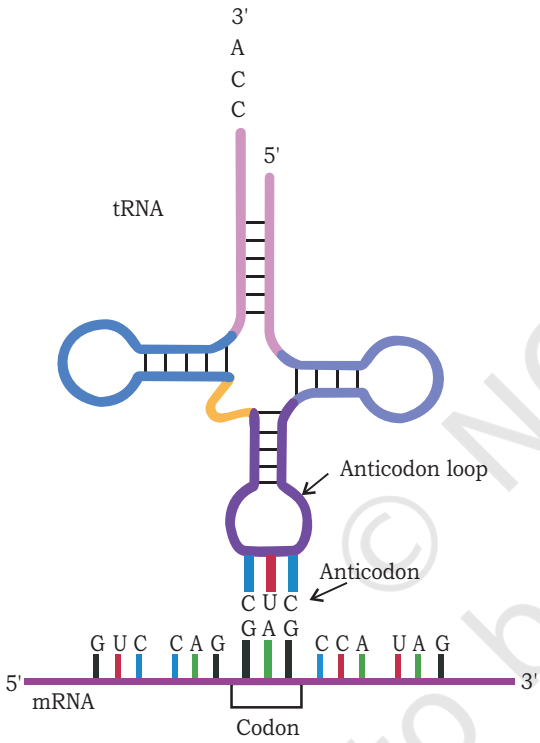


Fig. 7.26: Anticodon of tRNA recognises and base pairs with codon of mRNA

Charging of tRNA

The tRNA molecules deliver amino acids on to ribosomes. Three bases present in anticodon region of a tRNA recognise specific base of a codon of mRNA and get paired with them (Fig. 7.26). For example, the mRNA codon GAG contains information for glutamic acid. The anticodon of tRNA that base pairs with codon GAG by hydrogen bonding is CUC and carries glutamic acid at its other end (3' end). During translation, when an mRNA molecule moves through ribosome, glutamic acid is added to the polypeptide chain whenever GAG is presented for translation.

Each tRNA molecule carries a specific type of amino acid. The tRNA molecules are named according to the amino acid they carry. For example, the tRNA carrying methionine is called methionyl tRNA or $tRNA^{met}$. Similarly tRNA that carries serine is called $tRNA^{ser}$. In genetic code, 61 codons code for 20 amino acids. For 61 codons there should be 61 different tRNA molecules with different anticodon in the cell. However, the number of tRNA molecules is much less than 61. Hence, the anticodon of one tRNA molecule can recognise more than one codon on mRNA and base pairs with it. But how does the anticodon of one tRNA molecule recognise more than one codon? The base pairing between codon and anticodon follow Watson and Crick base pairing, i.e., A pairs with U and G with C. The pairing is precise in first two positions while it is flexible at third base of the

© NCERT not to be republished

codon. This unusual pairing at third base position is called **wobble pairing** (Wobble hypothesis) (Fig. 7.27).

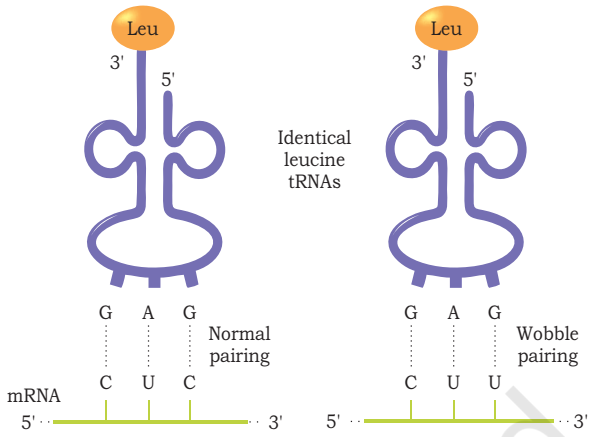


Fig. 7.27: Wobble pairing at third position of codon

Amino acids are first activated in the presence of ATP by aminoacyl-tRNA synthetase and transferred to its specific tRNA molecules (Fig. 7.28). A cell has 20 different aminoacyl-tRNA synthetase enzymes, one for each of the 20 amino acids. Each synthetase recognises a specific amino acid and transfers it to a specific tRNA molecule. Before starting of protein synthesis in ribosomes the amino acids are activated and transferred to their appropriate tRNA molecules. This is called charging of tRNA and involves two steps.

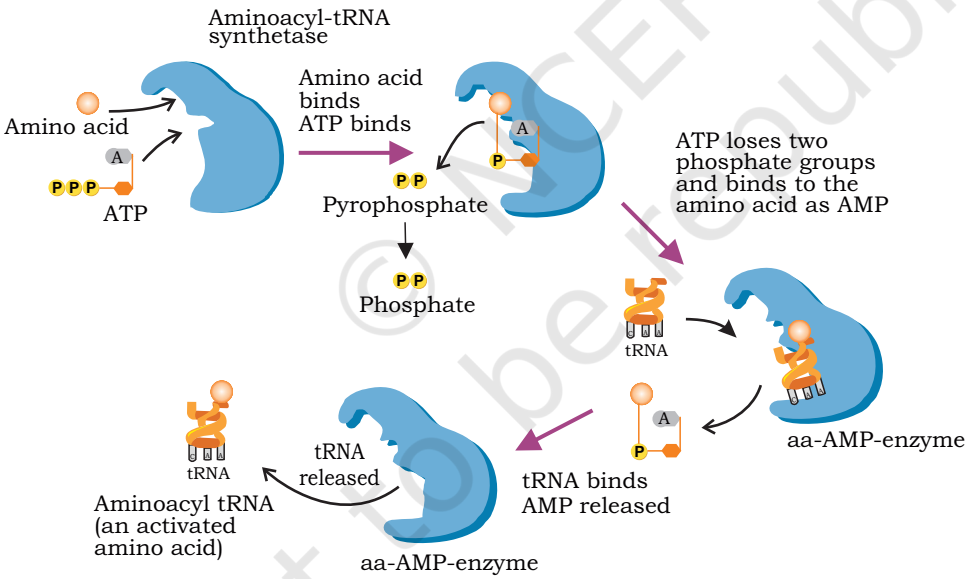


Fig. 7.28: Activation of amino acid and transfer to its specific tRNA to form aminoacyl-tRNA

In the first step, amino acid reacts with ATP in the presence of aminoacyl tRNA synthetase producing aminoacyl-AMP and PPi. In the second step, the amino acid is transferred to the tRNA. The activated amino acid is attached to the hydroxyl group of terminal adenine nucleotide present at the 3'OH end of the tRNA (CCA)

© NCERT not to be republished

sequence. The charged aminoacyl tRNA enters the ribosome.

$$\text{Amino acid} + \text{ATP} \xrightarrow[\text{tRNA synthetase}]{\text{Amino acyl}} \text{Aminoacyl-AMP} + \text{PPi}$$

$$\text{Aminoacyl-AMP} + \text{tRNA} \longrightarrow \text{Aminoacyl-tRNA} + \text{AMP}$$

Initiation of Translation

In the previous unit we have studied about prokaryotic (70S) and eukaryotic (80S) ribosomes and their subunits. There are three sites on a ribosome for binding of tRNA molecules: the A site (aminoacyl tRNA binding site), P site (peptidyl tRNA binding site) and E site (tRNA exit site). Aminoacyl-tRNA molecules enter into A site one after another during translation.

During initiation, the 30S subunit of ribosome in prokaryotes binds to ribosome binding sequence (Shine

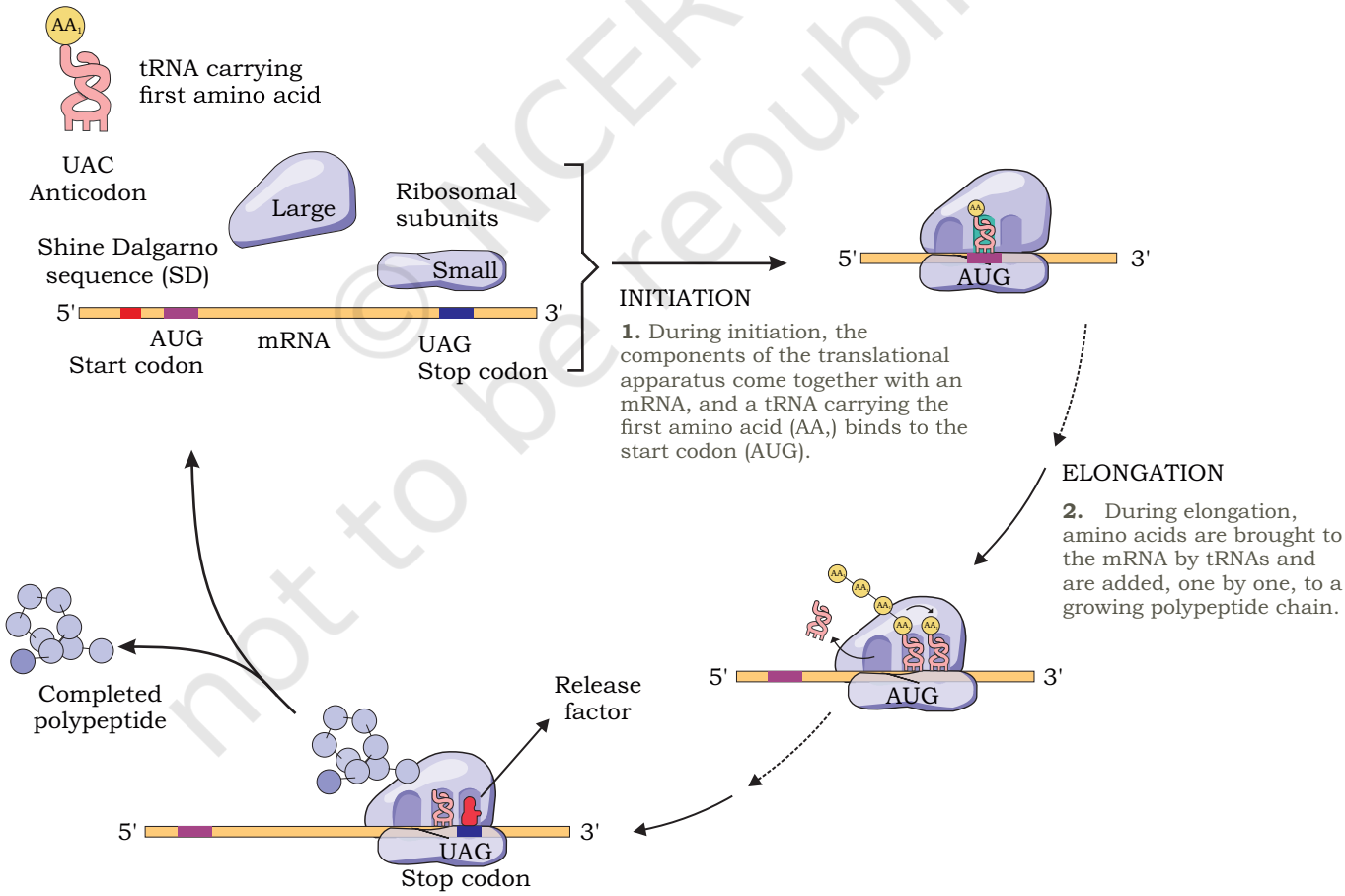


Fig. 7.29: Translation process in prokaryotes

© NCERT not to be republished

Dalgarno sequence) present at the 5' end of mRNA (Fig. 7.29). As a result, the initiation codon AUG (codes for methionine) is positioned in the P site.

The initiating amino acid methionine in prokaryotes is formylated (formylated methionine - fMet) but it is not formylated in eukaryotes. The initiating charged tRNA or aminoacyl-tRNA molecule in prokaryotes is fMet-tRNA^{fmet} (in eukaryotes—Met-tRNA^{met}) which attaches to the initiation codon AUG present in the P site. The 3' UAC 5' anticodon of fMet-tRNA^{fmet} base pairs with 5' AUG 3' codon of mRNA. The 50S subunit then associates with 30S subunit to form 70S initiation complex. The formation of this initiation complex also requires the activity of certain protein factors known as initiation factors and GTP. Hence, in the initiation complex, we have the initiator aminoacyl-tRNA (fMet-tRNA^{fmet}) in the P site, while the A site is empty, awaiting delivery of the second aminoacyl tRNA.

The initiation in eukaryotes is almost similar with some important differences. The small subunit of eukaryotic ribosome (40S) recognises the cap with the help of initiation factors, binds to it, and then moves along the mRNA until it locates the initiating AUG codon.

Elongation

The next step in protein synthesis is elongation, in which amino acids are joined to create a polypeptide chain. In the initiation complex, P site is occupied by amino acyl-tRNA with formylated methionine (fMet-tRNA^{fmet}) in prokaryotes and methionine (Met-tRNA^{met}) in eukaryotes (Fig. 7.30). The A site is vacant. Now a second aminoacyl-tRNA enters into A site with appropriate anticodon that base pairs with the mRNA codon. A peptide bond is now formed between the amino acids that are attached to tRNA at P and A sites. The peptide bond is formed between the carboxyl group of amino acid bound

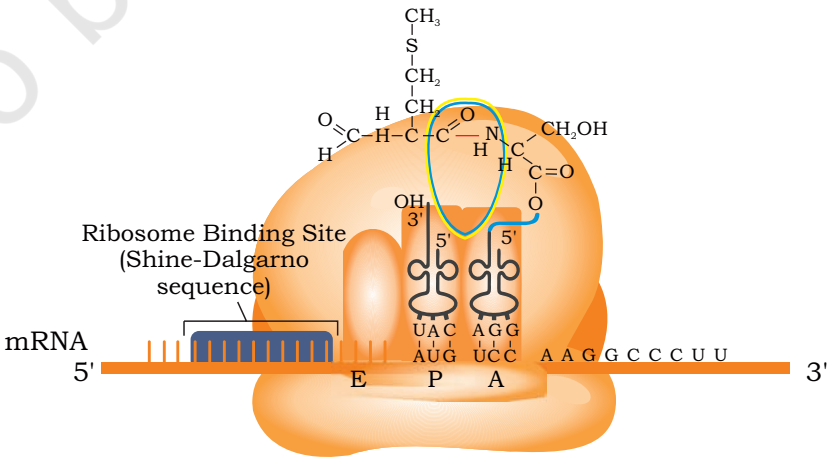


Fig. 7.30: Formation of peptide bond between initiating amino acid (fMet) and second amino acid

© NCERT not to be republished

to initiator tRNA in the P site and the free amino group of the amino acid attached to tRNA in the A site. This reaction is catalysed by peptidyl transferase enzyme. The formation of peptide bond releases the amino acid in the P site from its tRNA. Thus A site has a dipeptidyl-tRNA and P site contains an uncharged tRNA (without amino acid).

The ribosome now moves along mRNA in 5'→3' direction three nucleotides at a time. This movement is called translocation. The movement brings the uncharged tRNA from P site to E site from where it is ejected from ribosome. The peptidyl-tRNA with growing polypeptide chain moves from A site to P site. A site of ribosome is now again vacant with a new codon of mRNA. It is now ready to receive the next aminoacyl-tRNA molecule specified by the codon. The entire process is repeated and elongation of polypeptide chain takes place. Several protein factors called elongation factors and GTP are involved during the elongation step.

Termination

Elongation of polypeptide chain continues until a stop codon on the mRNA enters the A site of the ribosome. The three stop codons—UAA, UAG and UGA do not code for any amino acid. There are no tRNAs with anticodons complementary to these stop codons. No tRNA with amino acid enters into A site of ribosome when termination codon occupies it. The protein factors called release factors recognise stop codons and binds to A site (Fig. 7.31). The release factors then release the polypeptide chain from tRNA in the P site. Other protein factors bring about the release of tRNA from the P site, mRNA from ribosome and finally the dissociation of ribosome.

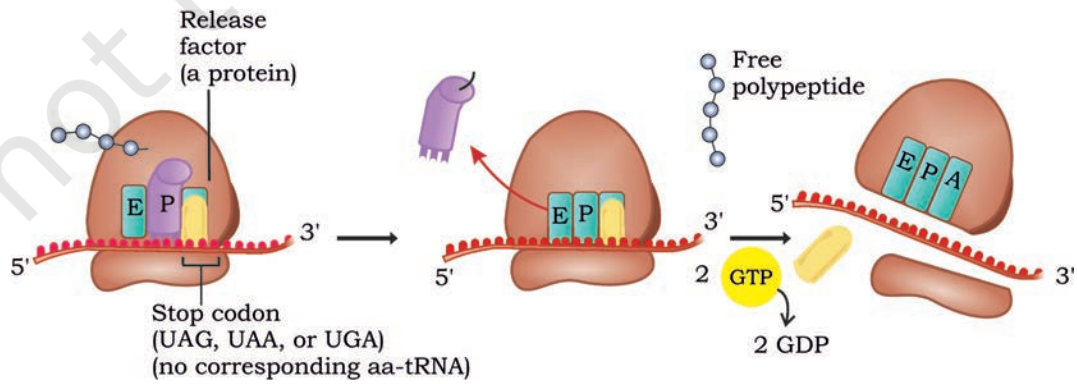


Fig. 7.31: Release factor recognises stop codon and terminates translation process

© NCERT not to be republished

Proteins synthesised in ribosomes then undergo **post-translational modifications** (PTMs) to form the mature protein product. Such modifications come in a wide variety of types, and are mostly catalysed by enzymes that recognise specific target sequences in specific proteins.

Polyribosomes

A single mRNA is translated simultaneously by several ribosomes producing many copies of a polypeptide chain. When the first ribosome attached to mRNA translocates far enough past the start codon, a second ribosome attaches to the same mRNA, eventually resulting in a number of ribosomes attached to mRNA called polyribosomes (Fig. 7.32). It is found in both prokaryotes and eukaryotes.

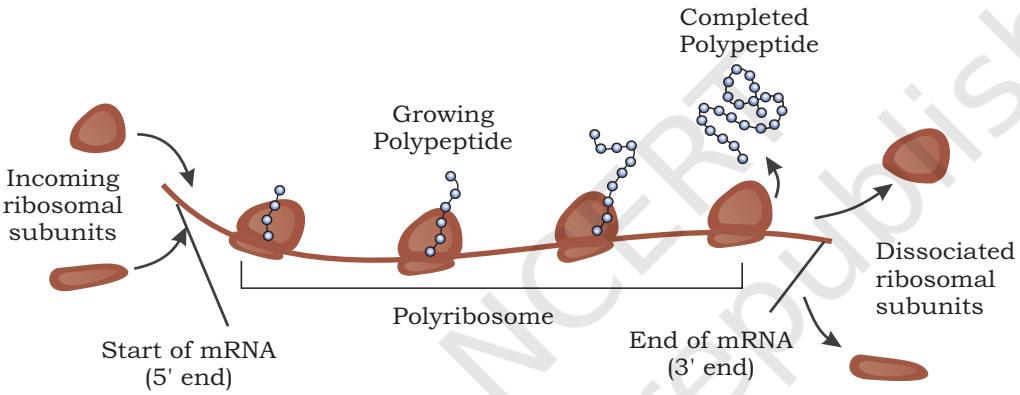


Fig. 7.32: Several ribosome bind to an mRNA during translation forming polyribosome

7.7 Gene Mutation

You have understood the fact that traits or characters in an organism are regulated or controlled by genes, which are a part of the DNA of the chromosomes. Traits are faithfully inherited in the form of coded information through these genes present on the DNA or chromosomes from parents to the offspring. All the mechanisms and processes such as DNA replication, transcription and the distribution of chromosomes during the process of cell division (mitosis or meiosis), etc. are extremely precise and accurate under the control of specific enzyme. Yet there are possibilities that some error may occur during these molecular processes leading to changes in the chromosomal organisation and molecular structure of the

© NCERT not to be republished

DNA carrying genes to be transmitted. These changes are broadly classified as **mutation** or sudden changes in the genetic material. Thus, it is now clear that changes may take place in the carrier of genetic information, i.e., DNA or chromosome both in somatic as well as germinal cell. However, such a change unless occurring in germinal cells and getting inherited into offspring has no significance.

Let us now examine different categories of changes or modifications that may happen. Among eukaryotes, chromosomes present in a germinal cell carries the hereditary information from parents to offspring. Hence, any change, either in the structure of a chromosome, (chromosomal aberration) or overall number of chromosome (ploidy) is categorised to be chromosomal mutation. Aberration may take place either due to loss of a part or addition of some part in a chromosome. Even rearrangement of chromosomal segment either within a chromosome or between two chromosomes are categorised as chromosomal aberration. Many extraneous factors like ionising radiations or some chemicals may induce such aberration in chromosomes called mutagen. You will appreciate that all such rearrangements can be identified by either specific chromosome staining technique called banding or fluorescence in situ hybridisation (FISH), which you will study later in Unit V. Similarly, there may be some exceptional situations in which the overall number of chromosomes (which remains constant from one generation to other) may get changed either by increase or decrease by one or both the homologues. Such a situation is described in the category of aneuploidy and you will find later that this is responsible for different types of hereditary syndromes observed in human beings. Likewise, change in number may occur by multiplication of the complete haploid set in such a way that the number increases to that of 3n, 4n or even more which is called polyploidy.

You may be thinking now that if such a change at chromosomal level may take place or be induced, there may be possibility that changes in the genetic material, i.e., DNA or RNA may also occur at molecular level. Bacteria has only a circular DNA or each of the chromosomes of a eukaryote contains DNA and all are involved in the process of making its copy by the process called replication before

© NCERT not to be republished

mitosis or meiosis. Any error happening during replication or by other means may alter the reading frame of the code of one or the other gene and may alter the genetic code and thus may affect the trait encoded by the gene. Such a change in the genetic material at molecular level is categorised to be gene mutation or point mutation. Sickle cell anaemia is one such example of point mutation in which substitution of one nucleotide results in the formation of abnormal sickle haemoglobin in human RBC and consequently the disease. Extraneous agent (mutagen), physical (ionising radiation, UV rays), chemical or biological (viruses) may induce gene mutations. We will focus our attention mainly on gene mutation.

It is now clear that alteration in the genetic material, i.e., DNA (RNA in case of a few viruses) may occur during the molecular processes. It has been observed that there are some intrinsic properties of the molecule or process that may result into changes in DNA or gene from the point of view of structural organisation at molecular level. These changes can be categorised in three different groups, i.e., addition, deletion and substitution of one or a few nucleotides. Among these, the two types of changes namely addition and deletion change the entire reading frame of nucleotide sequence on the DNA molecule. The impact of such a change can be understood by the fact that the changed coding of DNA may change its expression during RNA transcription and ultimately during the polypeptide chain synthesis. Obviously, the protein synthesized by such modified gene may not be a normal one or even the specific protein may not be synthesised.

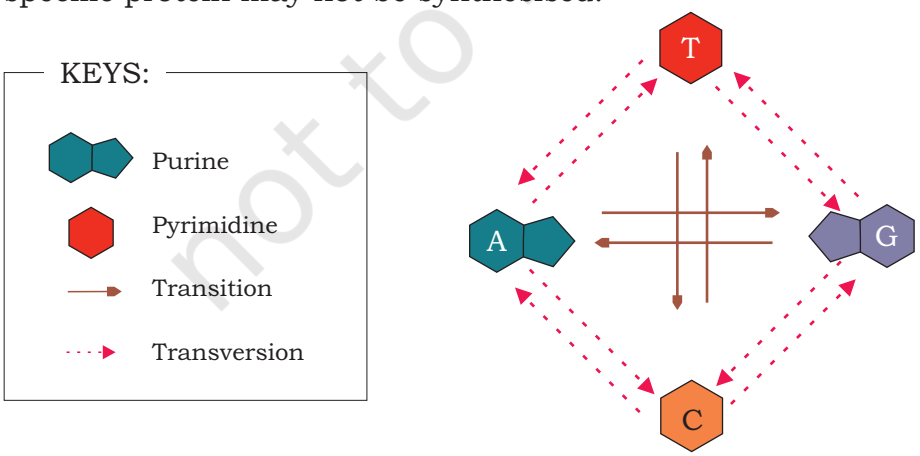


Fig.7.33: Different substitution mutations

© NCERT not to be republished

The change in the gene may also happen due to substitution or replacement of single nucleotide. The replacement of nucleotide may take place in which a purine base is replaced by another purine base. Same may happen for pyrimidine base too. Such a substitution type of mutation is known as **transition**. Similarly, a purine base may be replaced in a gene by a pyrimidine base or its *vice versa*, which is called **transversion** (Fig. 7.33).

7.7.1 Molecular mechanism of mutation

As discussed earlier, mutations may happen spontaneously due to the intrinsic properties of the molecule or they may even be induced by external agents called **mutagens**.

Spontaneous mutations—DNA carries the genetic information in the form of nucleotide. Among these nucleotides, there are functional groups present either in C=O or C–NH_2 form, normally called keto or amino form

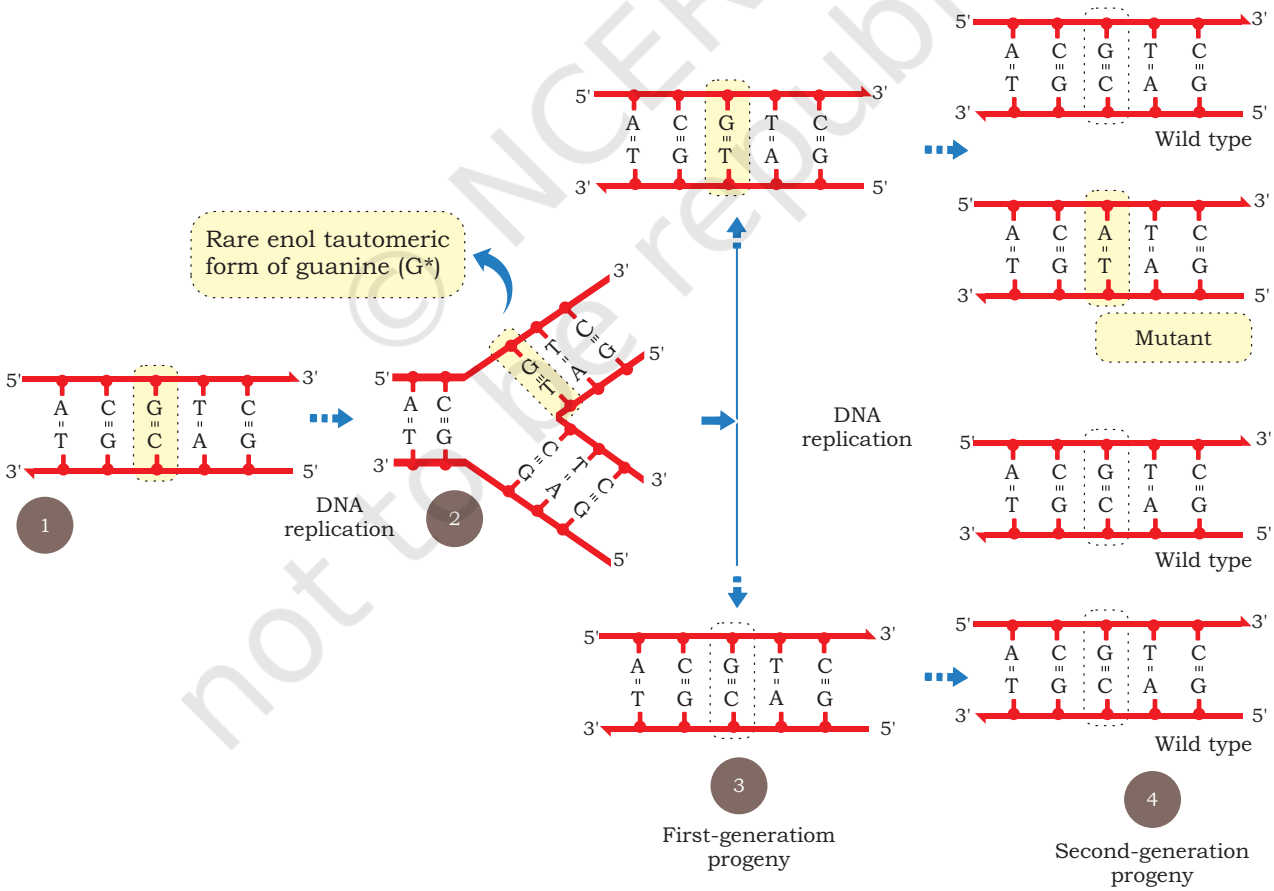


Fig. 7.34: Showing spontaneous induction of substitution mutation

© NCERT not to be republished

respectively. However, in these forms of nitrogenous bases hydrogen atom may shift from one atom of the molecule to the other atom within the base. Such a phenomenon is known to be the tautomeric shift and results in to a temporary phase of the nitrogen base called either enol (C-OH) or imino (C=NH) form. These rare tautomeric forms of nitrogenous base have changed property to pair with other nucleotide in the DNA molecule. Therefore, at the time of replication, when the imino form of guanine is present in the DNA, it makes a complementary pairing with thymine whereas the former would have normally paired with cytosine. During the next replication cycle, place where thymine is incorrectly added may normally pair cytosine leading to the substitution of G≡C pair in gene by A=T pair (Fig. 7.34). We have discussed the consequence of single base substitution in case of sickle cell anaemia.

Induced Mutation—It is after the use of Hermann J. Muller's experiments of inducing mutation using X-ray, a new area of induction of mutation using various external agents were opened. These mutagens fall in the categories of:

- **Physical**— Radiations like X-ray, UV rays, etc.
- **Chemical**— Alkylating agents like Mustard gas, ethyl methane sulfonate (EMS); base analogs like 5-bromouracil or 2-aminopurine; the deaminating agents like nitrous acid, etc.

In order to understand the mechanism of induction of mutation by external agent let us try with a few examples. Any cell, especially those responsible for the formation of gamete when gets exposed to ionising radiation of X-ray it may induce breakage of different bonds present within the molecule. In case such a breakage occur in the phosphodiester bond of DNA, it may result into the loss of a few nucleotides in it. This may result into the frame shift mutation due to deletion of one or a few nucleotides. Consequence of such a change in the reading frame of the genes is understandable as it may lead to a wrong RNA transcription and non-expression or altered expression of the gene. Even non-ionising radiation like UV rays may lead to the excitation of electron within the DNA molecule. These excited molecules may become more reactive and may lead to the induction of deletion or substitution type of mutation.

© NCERT not to be republished

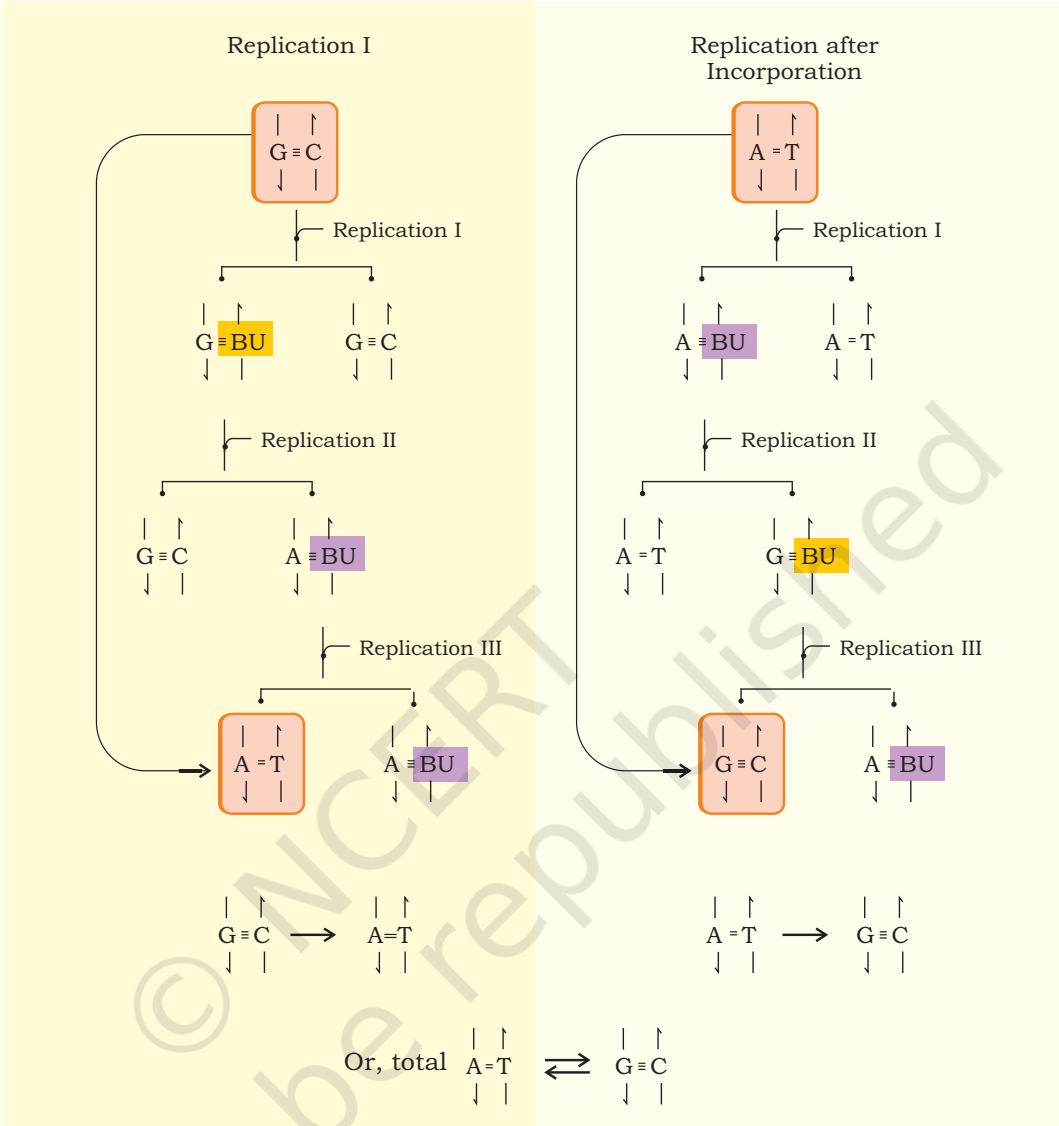


Fig. 7.35: Effect of 5-bromouracil on DNA replication

Similarly, base analog like 5-bromouracil when present in its keto form may get incorporated preferentially in the DNA as a complementary base against adenine. However, in the eventuality of its conversion into tautomeric enol form 5-bromouracil may pair with guanine nucleotide resulting into the substitution of A=T pair into G≡C pair (Fig. 7.35).

Alkylating agent like EMS ethylates DNA nucleotide either at nitrogen at 7th position or oxygen at 6th position. Such an alkylation alters the pairing property. For example,

© NCERT not to be republished

the ethylated guanine nucleotide pairs with thymine, which means a G≡C pair in DNA will be mutated to A=T pair. Many such chemical mutagens induce mutation by the process of altered pairing property.

7.8 DNA Repair

From the above account on mutation it may perhaps be understood that the rate of mutation either spontaneous or induced must be very high. But such a high rate of mutation is not observed. Also, a very high rate of mutation is not expected considering the fact that the genetic material gets transferred in a stable way. Understanding of the molecular processes in various organisms has revealed that mechanisms exist so that most of the errors that occur are corrected. Many of such mechanisms have been identified in the bacterial system *E. coli*. Out of these we will see a few to understand the mechanism.

Excision repair— This is a mechanism in which altered or modified bases in the DNA are removed or excised sequentially by identifying the altered base and subsequently removing them by enzymatic binding.

Gap thus created is eventually filled by a DNA polymerase. Following are the two important excision repair mechanism.

i. **Base Excision Repair (BER):** In BER, an altered or damaged base is replaced by a correct base. DNA nucleotides can be damaged by oxidation, alkylation and deamination that result in incorrect base pairing leading to mutations. The enzyme DNA glycosylase identifies the damaged bases and removes them. The gap region so created is termed as the AP site (apurinic or apyrimidine site). In the next step, an endonuclease cleaves this region to yield free 3'-OH and 5'-PO^4 group (Fig. 7.36a). DNA polymerase then adds the correct bases and through its exonuclease activity removes the old damaged bases. Finally, DNA ligase seals the remaining gap by formation of phosphodiester bond.

ii. **Nucleotide Excision Repair (NER):** In case of NER, a comparatively large portion of altered part of DNA is repaired. DNA is susceptible to radiations like

© NCERT not to be republished

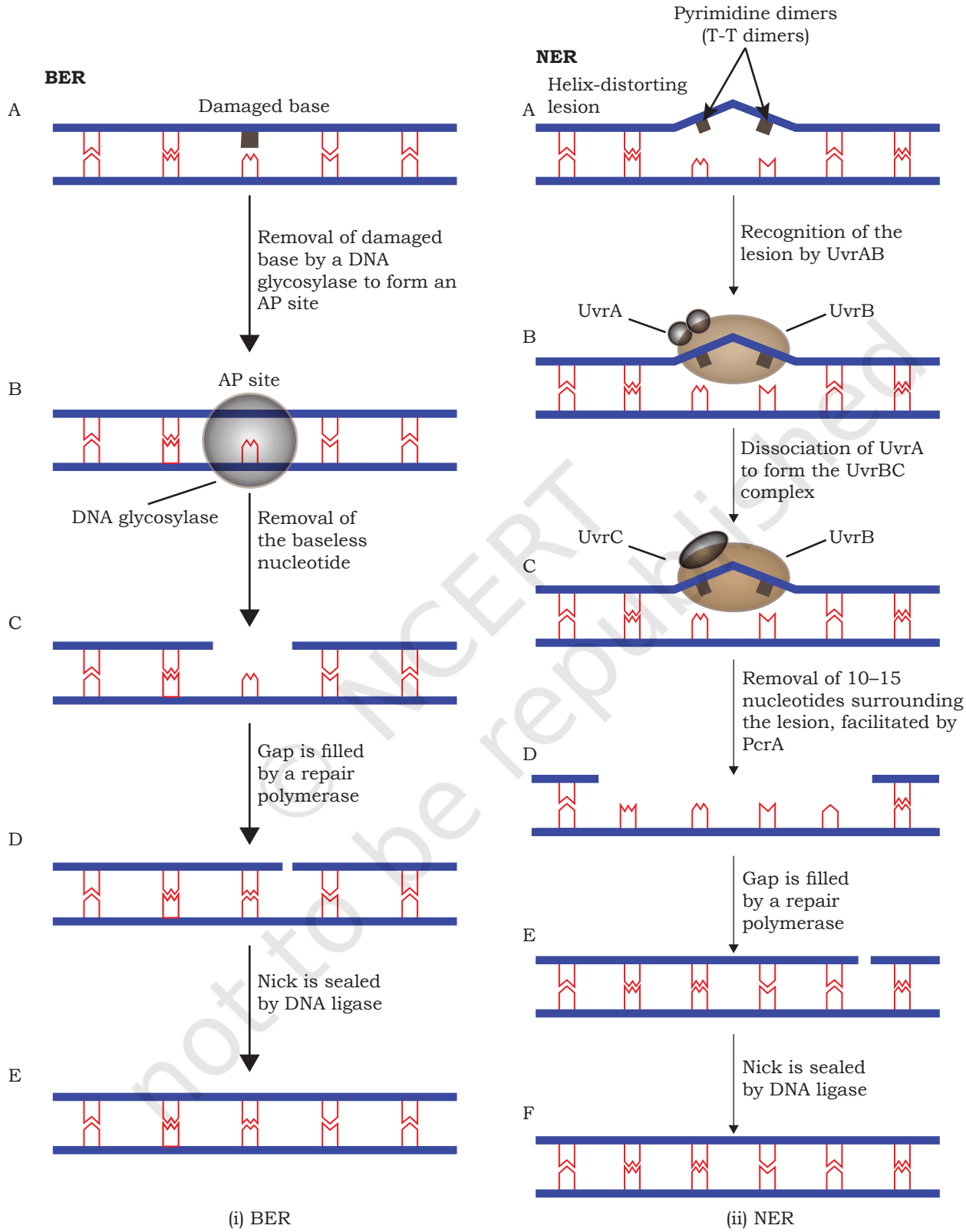


Fig. 7.36(a): (i) Base excision repair mechanism, (ii) N

© NCERT not to be republished

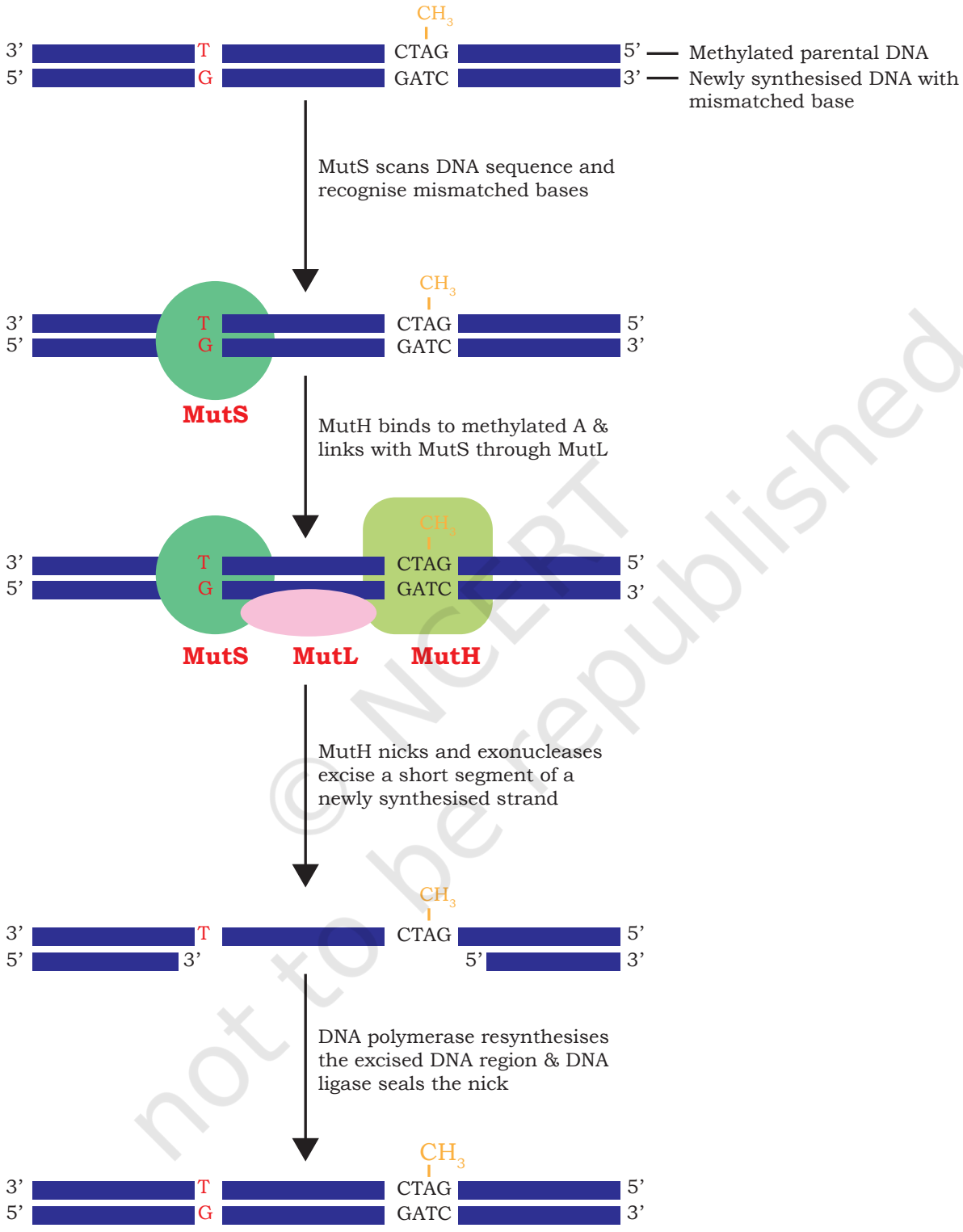


Fig. 7.36(b): Nucleotide excision repair

© NCERT not to be republished

UV radiation and it may lead to the formation of the pyrimidine dimers (T-T dimers), which distorts the DNA helix. There is a mechanism in the cell by which these dimers are removed. It is facilitated by a protein complex called UVr. (Fig.7.36b).

Initially a component of UVr complex scans the DNA for the presence of T-T dimers that distort the DNA helix. After identification, other components of UVr cleaves the phosphodiester bond four nucleotides downstream and eight nucleotides upstream to the DNA damage site. In the next step, the other component or UVr complex which has DNA helicase property unwinds the DNA and removes the cleaved nucleotide fragment. Finally, DNA polymerase I adds the nucleotide in 5'-3' region and the gap gets sealed by DNA ligase.

Mismatch repair (MMR)— During DNA replication, sometimes an incorrect nucleotide get incorporated. There exists a mechanism by which such a mismatch is corrected through different types of proteins namely MutH, MutL, MutS and MutT. These four proteins recognise and cleave the replicated DNA segment about 1,000 nucleotides from the point of mismatch. Gap thus created is filled by DNA polymerase followed by joining of nick by DNA ligase. This process is also important for proof-reading during replication.

A few other DNA repair mechanism have also been observed and studied both in prokaryotes and eukaryotes. You will study them in your higher classes. However, based on the account of mutation and repair mechanism the stable nature of genetic information can be appreciated vis-à-vis its ability to undergo changes.

7.9 REGULATION OF GENE EXPRESSION

Do you know in a multicellular organism there are many types of cells differing in structure and function, nevertheless their genes are identical? This is because all cells are derived from zygote by mitotic divisions. All activities of an organism are controlled by genes. Most of the genes of an organism express themselves by producing proteins. All genes are not expressed in all cells as their products are not needed at one time. Only those genes whose products are required in a cell are expressed while other genes are not expressed as its products are not

© NCERT not to be republished

required by the cell at that time. This 'switching on and switching off' mechanism of gene action is known as regulation of gene expression or regulation of gene action.

Lower organisms like bacteria encounter wide range of environmental conditions. For example, *E. coli* live in our large intestine. Our eating habits determine the nutrients available to this bacterium. When glucose is available, it expresses those genes whose products (enzymes) will break it down for generation of energy. If lactose or any other sugar is available then some other genes are expressed whose products will break it to generate energy. This indicates that specific genes are expressed at specific times according to the need of the cell.

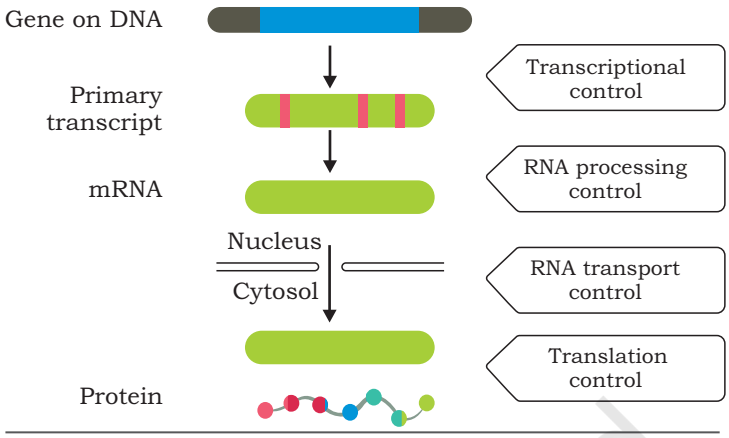


Fig. 7.37 Levels of regulation of gene expression

The expression of genes may be regulated at different steps along the pathway of flow of information from genotype to phenotype. Regulation may be at chromatin level, transcription level, mRNA processing (eukaryotes), transport of mRNA and at translational level (Fig. 7.37). In both prokaryotic and eukaryotic systems, transcription initiation is an important point of gene regulation as the cell not only decides which gene has to be expressed but also their degree of expression. In most cells of a species or an organism, some genes are expressed at a more or less constant level and are called **'housekeeping genes'** or **'constitutive genes'**. The product of these genes is required all the time. Genes encoding the enzymes that catalyse the steps in central metabolic pathways, such as the citric acid cycle fall into this category. Such unregulated expression of genes is called **constitutive gene expression**. But rate of expression of most of the genes alter in cells according to the molecular signal it receives. The product level of these genes rise and fall according to need of the cell. Such type of control is called **regulated gene expression**.

7.9.1 Regulation of gene expression in bacteria

The mechanism of regulation of gene expression was first studied in bacterial cells. The organisation of functionally

© NCERT not to be republished

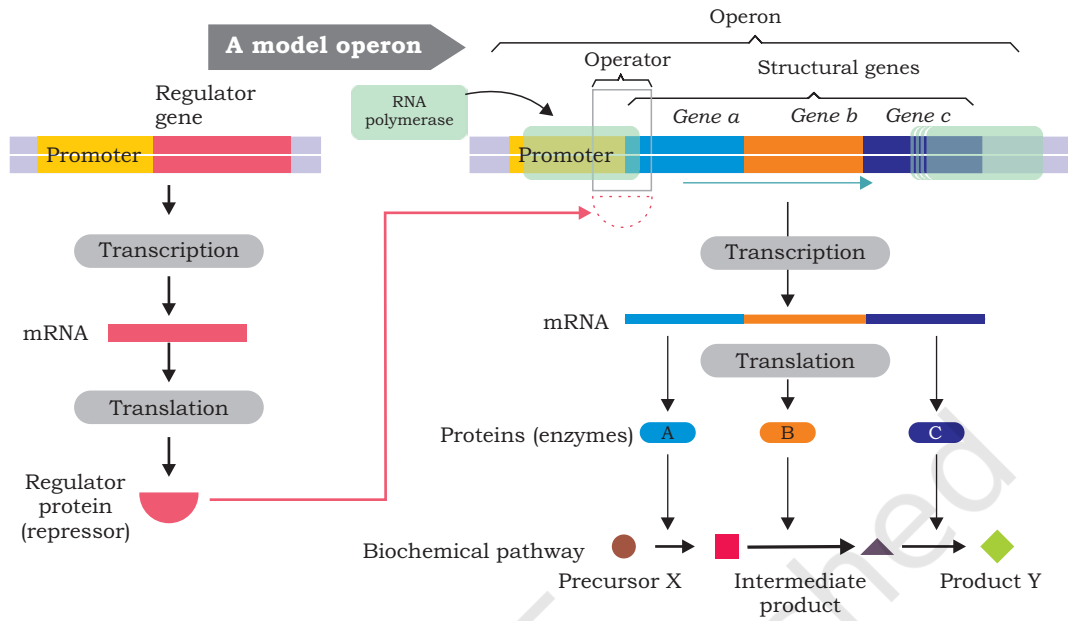


Fig. 7.38: Organisation of an operon

related genes in prokaryotes is different from that of eukaryotes. In bacteria, genes that have related functions are clustered and often transcribed together into a single mRNA molecule. The advantage of clustering related genes is that a single 'on-off switch' can control all the genes of a cluster. It means all the genes of a cluster are coordinately controlled. On the other hand, each gene of eukaryote is transcribed into a separate mRNA. A group of clustered structural genes that are transcribed together along with promoter and additional controlling sequences (that control transcription) is called an **'operon'**.

A typical operon (Fig. 7.38) has a set of structural genes or cistrons (encode proteins involved in metabolism) at one end which are transcribed and then translated to produce different proteins. Upstream to the first structural gene is **'promoter'** which controls the transcription of structural genes. RNA polymerase binding site lies in the promoter. A DNA segment called **'operator'** positioned within the promoter or between the promoter and the first structural gene controls the access of RNA polymerase to the genes. The operator is the site of binding of the repressor protein, the latter binds to the operator forming an operator-repressor complex. When the repressor binds to the operator, transcription of the structural genes cannot occur. A

© NCERT not to be republished

'regulator gene' is located upstream to the promoter of an operon. It is not considered part of the operon although it regulates the transcription of structural genes. It has its own promoter and is transcribed to produce a small mRNA which is then translated into a protein called regulatory protein **(repressor)**. Repressor may be either an active repressor or an inactive repressor. The active repressor protein binds to operator of the operon and prevents the binding of RNA polymerase to the promoter thereby interferes with the transcription of structural genes.

The mechanism of regulation of operon was first described by Francois Jacob and Jacque Monod in 1961 and proposed 'operon model' for the genetic control of lactose metabolism in *E. coli* cells. There are two types of operon, **inducible operon** and **repressible operon** based on the nature of their response to an effector molecule. In case of inducible operons, the effector molecules are called **inducers** (substrates) when present bind to active repressor and inactivates it. The inactive repressor-inducer complex cannot bind to operator, and transcription of structural genes in the operon is turned on (induced) and subsequently proteins (enzymes) are translated. The enzymes whose production can be increased by the presence of the substrate on which it acts are called **inducible enzymes** and the genetic system responsible for the synthesis of such an enzyme is called **inducible system**. In case of repressible operons, the effector molecules are called **co-repressors** (end products). When co-repressors bind to inactive repressors, the repressor-corepressor complex is active, bind to operators and prevent RNA polymerase from transcribing structural genes. For instance, when no amino acids are supplied from outside, the *E. coli* cells can synthesise all the enzymes needed for the synthesis of different amino acids. However, if a particular amino acid, for instance, histidine, is added, the production of histidine synthesising enzyme falls. In such a system, the addition of the end product checks the synthesis of the enzymes needed for the biosynthesis. Such enzymes whose synthesis can be checked by the addition of the end product are **repressible enzymes** and the genetic system is known as **repressible system**. There are two types of transcriptional control: **negative and positive control**. In

© NCERT not to be republished

negative control, the regulatory protein is a repressor that inhibits transcription. In positive control, the regulator protein is an activator which stimulates transcription.

7.9.2 The *Lac* operon – an inducible operon

The lactose (milk sugar, disaccharide) is a β-galactoside is available to *E. coli* in the colon when a person drinks milk. The bacteria uses lactose for energy as well as source of carbon after it is broken down into glucose and galactose by

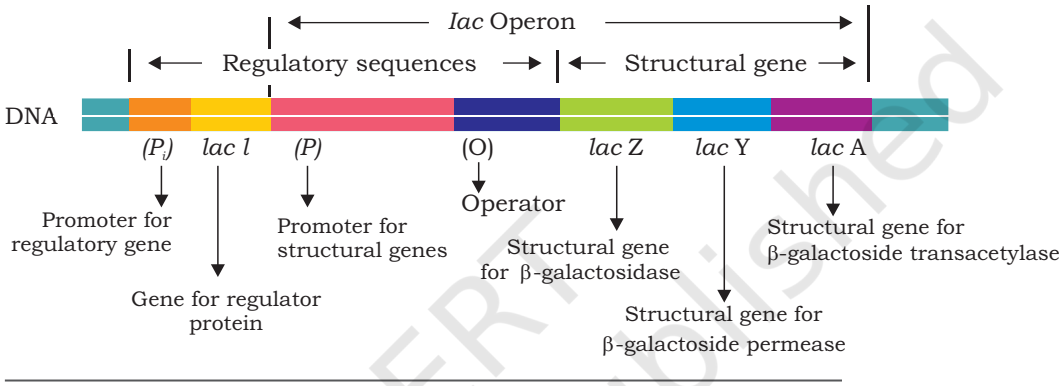


Fig.7.39: Structure of lac operon

enzyme β-galactosidase. When *E. coli* are growing in absence of lactose, few molecules of β-galactosidase are present in the cells but when lactose is added to the bacterium's environment, the number of β-galactosidase molecules in the cells increases many folds within 2 to 3 minutes.

The *lac* operon consists of three structural genes; *lacZ*, *lacY* and *lacA* encoding three different proteins (Fig. 7.39). The *lacZ* gene encodes **β-galactosidase** that break down lactose into glucose and galactose. This enzyme can also convert lactose into allolactose which act as an inducer in *lac* operon. The gene *lacY* encodes **β-galactoside permease**, a membrane protein which actively transports lactose into the cell. The *lacA* encodes **β-galactoside transacetylase**, but its function in lactose metabolism is not yet known.

The regulator gene called *lacI* is located upstream to promoter of *lac* operon has its own promoter. It is transcribed into a small mRNA which is then translated into a regulator protein called repressor protein. The repressor is an allosteric protein has two binding sites; one for binding to operator of operon and to the other binds inducer (allolactose).

© NCERT not to be republished

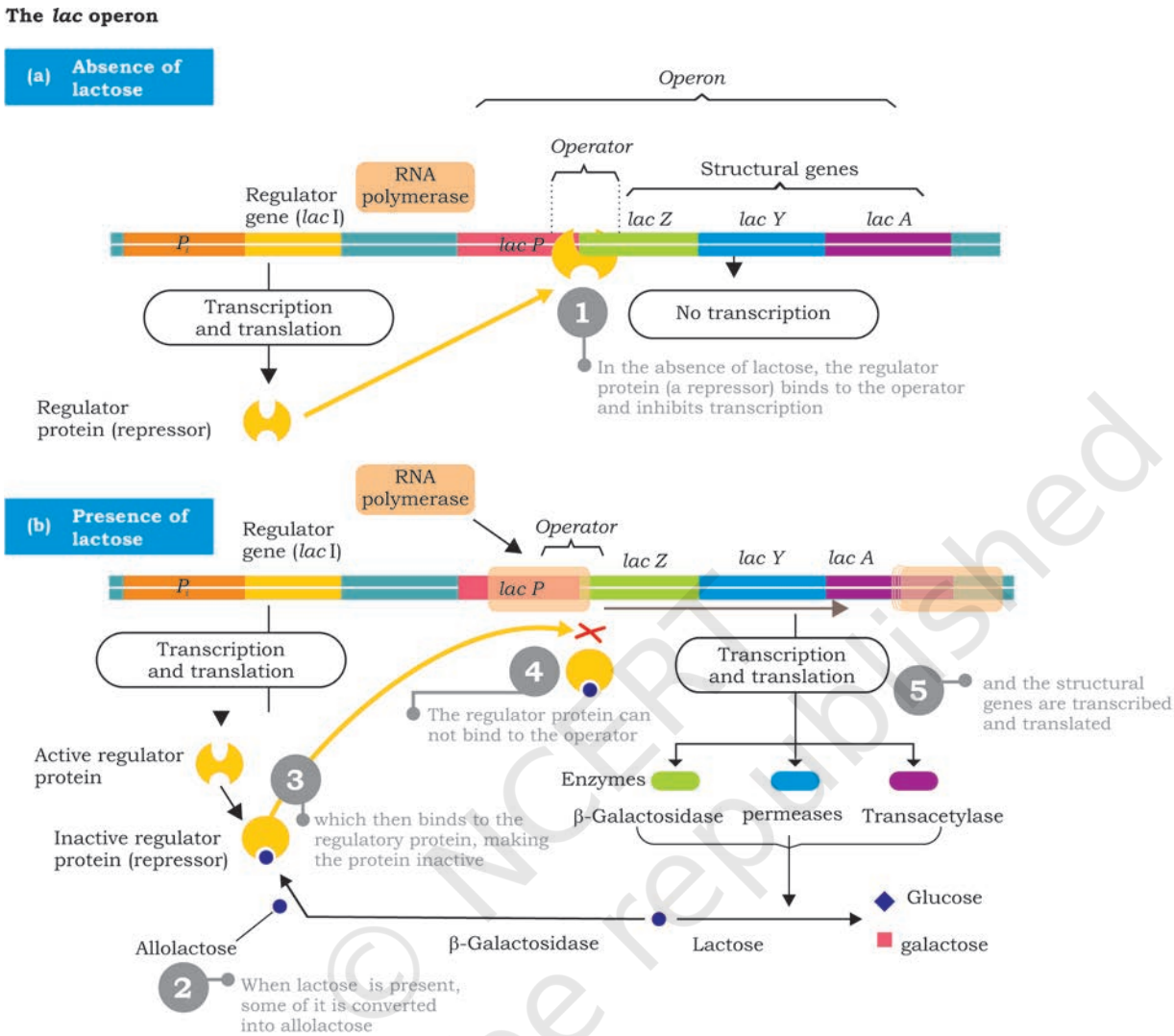


Fig. 7.40: Regulation of lac operon

In the absence of lactose in *E. coli* cells, the repressor protein encoded by *lacI* is active and binds to *lac* operator; physically blocking the binding of RNA polymerase and transcription of structural genes is prevented (Fig. 7.40). This is negative control of lac operon. As long as the repressor is binding with the operator, no proteins are made. However, when lactose is present, β-galactosidase converts some of it into allolactose. The allolactose acts as inducer, binds to active repressor and causes conformational change by which it becomes inactive. The inactive repressor fails to bind with operator and binding of RNA polymerase is no longer blocked. RNA polymerase now transcribes *lacZ, lacY* and *lacA* into a polycistronic

© NCERT not to be republished

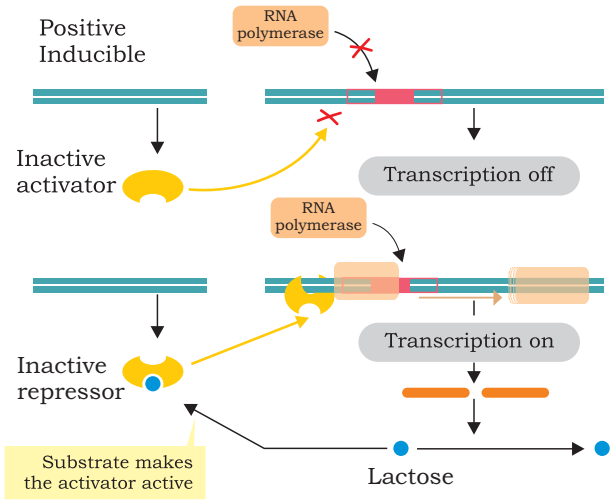


Fig. 7.41: Positive control of lac operon

mRNA which translates into three different enzymes required for lactose metabolism (Fig. 7.41). The production of the enzymes to break down lactose continues until enough of the lactose molecules are broken down and then release repressors to recombine with the operator to stop production of the enzymes. The *lac* operon is an inducible operon as presence of lactose induces the production of β-galactosidase, β-galactoside permease and β-galactoside transacetylase.

In positive control of *lac* operon, the regulator protein, i.e., an activator binds to DNA at a site other than operator. The activator is produced in an inactive state and fails to bind to DNA (Fig. 7.41). The RNA polymerase does not bind to promoter and transcription is off. When inducer associates with inactive activator rendering it active, RNA polymerase binds to promoter and initiates transcription.

SUMMARY

- DNA was, for the first time, isolated from nuclei of pus cells by Johann Friedrich Miescher in 1869.
- The phenomenon of transfer of genetic material from one cell to another that alter the genetic make-up of the recipient cell is called transformation and this was discovered by Frederick Griffith in 1928.
- The experiments conducted by Oswald Avery, Colin Macleod and Maclyn McCarty revealed DNA as the likely transforming agent.
- The experiments conducted by Hershey and Chase in 1952 provided strong evidence that DNA is the genetic material.
- Gene is the unit of inheritance that controls a specific trait or character and may also be expressed in alternative forms known as alleles.
- The expression of DNA through the synthesis of polypeptide chain via RNA synthesis represents the Central Dogma of genetics.
- The fact that one gene encodes one polypeptide, the central dogma also got modified from one gene, one protein to one gene, one polypeptide.

© NCERT not to be republished

- Each gene must satisfy the test to be a unit of function, unit of recombination and unit of mutation.
- The kind of DNA replication in which the parental duplex DNA form two identical daughter duplex, each of which consists of one parental stand and one newly synthesised daughter strand is called semi-conservative replication.
- It was Messelson and Stahl who experimentally distinguished between the old and new stand of the DNA after replication using two isotopes of Nitrogen, ^{14}N and ^{15}N in their experiment.
- Several enzymes and proteins are involved in the replication of DNA in both prokaryotes and eukaryotes such as DNA polymerase, primase, helicase, single-strand binding protein, topoisomerase, DNA ligase.
- Replication usually starts at a specific site on a DNA sequence known as origin of replication.
- The new strand that is synthesised continuously in 5'→3' direction is called the leading strand.
- The DNA strands synthesised discontinuously with the Okazaki fragments are called the lagging strand.
- Transcription is a process by which genetic information is transferred from DNA to RNA.
- Translation is the process by which information is transferred from RNA to polypeptide chain and consists of four stages such as (i) charging of tRNA (ii) initiation (iii) elongation, and (iv) termination.
- In case of reverse transcription the information present in the genetic material which is RNA is first transferred to a single stranded complementary DNA strand and is then converted into a double stranded DNA.
- Within a gene, only one of the nucleotide strands is normally transcribed into RNA. The DNA strand whose nucleotide sequence is complementary to that of the mRNA is called the template or antisense strand, while the other strand whose base sequence is identical to that of the mRNA (except for T in DNA and U in RNA) is called the sense or coding strand.
- The major steps which are common to both prokaryotes and eukaryotes in the process of transcription include initiation, elongation and termination. In addition to these steps, in eukaryotes the primary transcripts undergo post-transcriptional modifications such as capping, splicing, poly-adenylation.
- Genetic codes are triplet codons, i.e., unique combinations

© NCERT not to be republished

of three bases which codes for a specific amino acid depending on their combinations. There are 64 codons, out of which 61 codons code for 20 amino acids.

- Polyribosomes consist of several ribosomes attached to the same mRNA.
- Mutation is the alteration in the genetic material, i.e., DNA (RNA in case of a few viruses) and can be categorised as addition, deletion and substitution of one or a few nucleotide.
- High rates of mutation are not observed due to the DNA repair mechanism which could be through excision repair, mismatch repair, etc.

EXERCISES

1. What is the importance of gene expression? What are the steps involved in it?
2. Describe the process of regulation of gene expression in prokaryotes by giving example of *lac* operon.
3. What would be the effect of loss of all proteins from a cell on DNA replication?
4. How is the structure of DNA affected by UV rays? Discuss the molecular basis of the type of mutation caused by this type of radiation and the mechanism used by cells to correct them.
5. Differentiate between the following
   (a) Leading strand and lagging strand
   (b) Transcription and translation
   (c) Transition and transversion mutation
   (d) Codon and anticodon
6. Which of the following types of radiations is least likely to be harmful to cells?
   (a) Gamma rays
   (b) Ultraviolet rays
   (c) X rays
   (d) Alpha rays
7. In which of the following DNA repair mechanism is apyrimidinic or apurinic (AP) site formed?
   (a) Excision repair
   (b) Mismatch repair
   (c) Both of the above
   (d) None of the above

© NCERT not to be republished

11150CH08

Chapter 8

Genetic Disorder

8.1 *Chromosomal Abnormalities and Syndromes*

8.2 *Monogenic Disorders and Pedigree Mapping*

8.3 *Polygenic Disorders*

8.1 Chromosomal Abnormalities and Syndromes

In certain situations e.g., due to environmental radiation, food intake or internal genetic conditions, chromosomes may suffer damage or may change in numbers. The change in structure is called **structural chromosomal abnormality (or aberration)** and the change in number is called **numerical chromosomal abnormalities**. When one chromosome of the pair is absent, the condition is called monosomy (2n-1) for that chromosome e.g., monosomy of chromosome 1. When a chromosome is present in three copies, this condition is called trisomy (2n+1) e.g., trisomy of chromosome X. It is important to notice that both monosomy and trisomy come under the broad category of aneuploidy. However, when the entire set of chromosome is multiplied (e.g., 69: 23 × 3, 92: 23 × 4), the condition is called polyploidy. The artificial breeding of plants has resulted in several polyploid varieties that we commonly use in our food. For example, bread wheat has six sets of chromosomes (hexaploid), cabbages or mustards are

© NCERT not to be republished

tetraploids. Likewise, banana and apple are triploid (3 sets of chromosomes), strawberry and sugar cane are octoploid (8 sets of chromosomes). Both structural or numerical changes can result in significant changes in phenotypic condition in the form of diseases or syndromes.

8.1.1 Structural chromosomal abnormalities

Structural chromosomal abnormalities may be of following types:

1. **Deletion**— In deletion, a segment of a chromosome breaks away leading to shortening of the chromosome [Fig. 8.1(a)]. For example, retinoblastoma is caused due to deletion of a portion of chromosome 13. Sometimes when two ends of a chromosome are deleted, they can reattach to form a ring chromosome.

2. **Duplication**— Duplication refers to when a segment of the chromosome gets repeated resulting in a longer chromosome [(Fig. 8.1(b)]. This can lead to conditions e.g., Charcot-Marie-Tooth disease caused due to duplication of genes on chromosome 17.

3. **Inversion**— In inversion, a segment of the chromosome breaks away, completely reverses itself and reattaches with the chromosome. Here the overall length of the chromosome remains same but the orientation of genes is reversed by 180 degrees [(Fig. 8.1(c)]. For example, RCAD syndrome caused by inversion of a segment of chromosome 17.

4. **Translocation**— In translocation, a segment of a chromosome breaks away and reattaches itself with another chromosome. If there is a mutual exchange of segments between two chromosomes, it is called

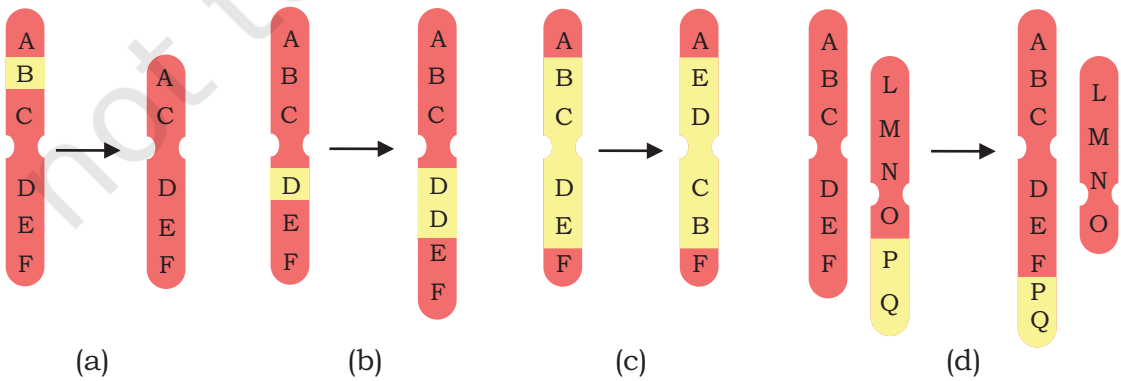


Fig. 8.1: (a) Deletion (b) Duplication (c) Inversion and (d) Translocation

© NCERT not to be republished

reciprocal translocation. Example: Burkitt's lymphoma, where exchange of materials happens between chromosomes 8 and 14. If a segment of a chromosome breaks away and attaches with another chromosome, without mutual exchange, it is called **Robertsonian translocation**. This may result in decrease of chromosome number of the cell [Fig. 8.1(d)].

8.1.2 Numerical chromosomal abnormalities

Some commonly observed syndromes/diseases due to numerical chromosomal abberations are described in following section. The term syndrome is generally referred to a group of symptoms which consistently occur together, or a condition characterised by a set of associated symptoms. A disease refers to abnormal physiological response to internal or external factors e.g., fever caused due to microbes.

1. Down's Syndrome

Incidence: Occurs in approx.1 per 800 live births.
Chromosomal basis: Down syndrome is a genetic condition that arises due to presence of an extra chromosome 21. Here, chromosome 21 is repeated thrice (trisomy 21), instead of showing up twice in a normal individual. The karyotype of Down syndrome is represented as 47, XX, +21 (females) and 47, XY, +21 (males) [(Fig. 8.2(a)].

Fig. 8.2: Karyogram of (a) an individual affected with down syndrome (b) an individual affected with Klinefelter's

The trisomic condition is usually caused by an error in the process of cell division called non disjunction, i.e., inability of chromosomes to separate at the time of cell division.

© NCERT not to be republished

The possibility of having a Down's syndrome baby in the family increases with the maternal age. It has been reported that more than 85% Down syndrome babies are born in mothers over 35 years of age, at the time of pregnancy.

Clinical symptoms: Some of the distinguishing features of Down's syndrome are: flat face, slanting eye, small mouth, protruding tongue, flattened nose, short neck, short arms and legs, single deep crease across the palm, low IQ, stunted growth, muscular hypotonia, under developed gonads. Down's syndrome babies also show breathing, heart or hearing problems.

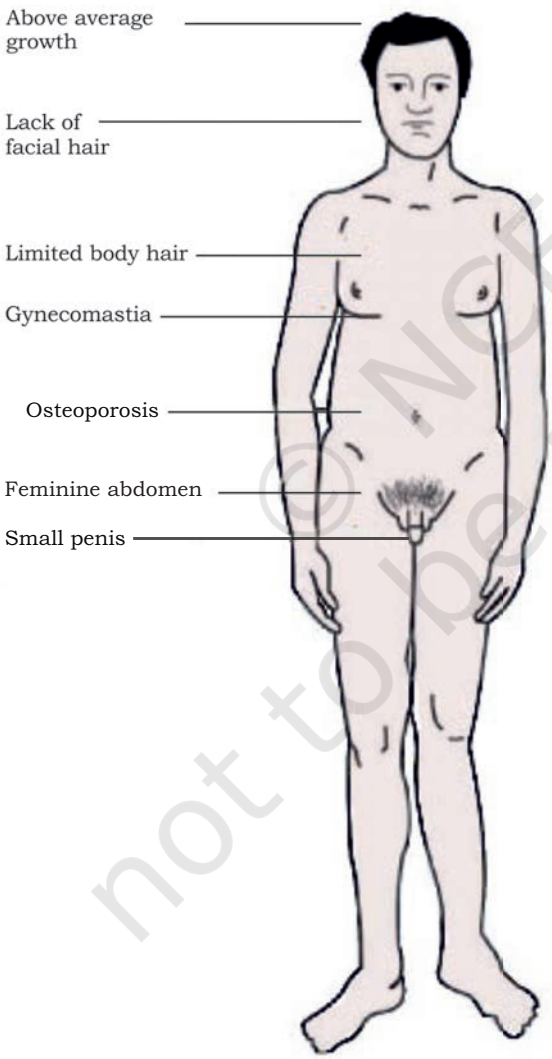


Fig. 8.3: Individual with Klinefelter's Syndrome

Diagnosis and Treatment: Down syndromes are usually diagnosed by an extra chromosome 21 in the karyotype. There is no single standard treatment protocol for Down syndrome. Treatments are tailored on specific set of conditions presented by these individuals. At early age, children with Down's syndrome can benefit from speech therapy, physiotherapy and taking nutritional supplements.

In early 1900s, on an average, Down's syndromes used to live until age 9. Now with the advances in diagnostic and treatment technologies, the age expectancy has increased up to 60 and even longer.

2. Klinefelter's syndrome

Incidence: Occurs in approximately 1 out of 1000 new born males.

Chromosomal basis: Genotype: 47, XXY. Affects males. The extra chromosome is not transmitted genetically (i.e., a Klinefelter newborn cannot have a Klinefelter father) but arises from inability of X chromosome to detach itself from the pair during meiosis (at the time of gamete formation). Fertilisation of an XX ova with a Y sperm produces an XXY zygote.

© NCERT not to be republished

Clinical Symptoms: Klinefelter's syndrome children are unusually tall for their age, have reduced facial and body hair, smaller testes, enlarged breasts and coarse voice [Figs. 8.2(b) and 8.3].

Diagnosis and Treatment: One of the most frequent methods to diagnose Klinefelter syndrome is through Barr body test of buccal smear. Normally no Barr body appears in the male buccal smear. However, in Klinefelter one Barr body shows up, indicating the presence of an extra X chromosome.

At the time of birth, babies with Klinefelter differ a little from other normal babies. However, as the age increase the differences become noticeable, especially at the time of puberty.

People with Klinefelter's syndromes are often treated with testosterone to look masculine. They also need to be psychologically counselled to control depression leading to aggression.

3. Turner's Syndrome

Incidence: Occurs in 1 in 2,500 newborn girls, frequently observed in miscarriages and still births.

Chromosomal basis: Affects females, arises due to the missing X chromosome in affected females. This is called monosomy X and the karyotype is represented as: 45, X. A cell division error during meiosis of an ovum results in an ovum with no X chromosome and other with two X chromosomes. The ovum with no X chromosome fuses with sperm with one X chromosome to generate 45, X condition. Mothers with Turner syndrome cannot pass the condition to their daughters i.e., this condition is not inherited.

Fig. 8.4: Turner's Syndrome

Clinical symptoms: Turner syndromes are diagnosed by following features — short stature, webbed neck (i.e., the neck skin is unusually loose and can be pulled several centimetres of the neck), small breasts, low set ears (i.e., ears are placed below the normal position), swollen hands and feet. Furthermore, ovaries are under developed and menstrual periods are usually absent (Fig. 8.4).

© NCERT not to be republished

Diagnosis and treatment: Prenatal chromosomal diagnosis usually happens through amniocentesis or chorionic villus sampling. At puberty the first test is that of Barr body of buccal smear. The absence of Barr body is the first indicator to follow up the condition with a detailed investigation. As with other syndromes, there is no permanent cure. However, growth and ovarian functions can be strengthened by the controlled administration of hormones androgen and estrogen.

Box 1

Famous people with syndromes

1. Isabelle Springmühl is a famous fashion designer with Down syndrome. Achieving this feat wasn't easy as several universities rejected her application to study fashion design. But she persisted and has showcased her talent in London, Rome and Mexico.
2 It's commonly believed that the famous US president, six feet two inches tall, Mr. George Washington had Klinefelter Syndrome. He had no children and adopted two later in his life.
3. Lauren Foster, a South African model was born male and diagnosed with XXY Klinefelter condition. However, Lauren chose to identify herself with feminine features and transitioned to a full female phenotype in her teens. Lauren became a successful model and appeared in Vogue Magazine. She even wanted to contest for the Miss South Africa pageant but was disqualified.
4. Hollywood TV, Film and Stage actress Linda Hunt was diagnosed with Turner syndrome. Linda started her career as a singer and made her Hollywood debut with Popeye film version. Linda has won 13 awards that include the 2012 Teen Choice Award and the 1984 Oscar Award for the Best Supporting Actress.
5. Dr. Catherine Ward Melver is a famous medical genetics doctor in the US. The 4 feet and 8 inch tall, Dr.Melver was diagnosed with Turner syndrome when she was seven years of age. Dr.Melver adopted a 4-year-old Turner syndrome girl, Zoe, from China.

8.2 MONOGENIC DISORDERS AND PEDIGREE MAPPING (CYSTIC FIBROSIS, SICKLE CELL ANEMIA, HAEMOPHILIA, COLOR BLINDNESS)

Monogenic disease is caused by an error in a single gene. According to current estimate over 10,000 of human diseases are estimatd to be monogenic affecting millions of individual world wide. The nature of disease, its sign and symptoms depend on the functions performed by the

© NCERT not to be republished

modified or defective gene. These diseases are inherited according to Mendel's Laws. In some cases, the mutation can be spontaneous and where we will not get the previous family history. There may be single mutation in one gene causing specific disease like sickle cell anemia or multiple types of mutation in one gene and producing the same disease like cystic fibrosis (more than 200 different types mutation can occur in one gene).

The single-gene or monogenic diseases can be classified into the following categories according to the inheritance pattern as follows:

- Autosomal recessive
- Autosomal dominant
- X-linked recessive
- X-linked dominant

For diagnosis of the inherited genetic disease one has to understand the concept of pedigree analysis. The pedigree analysis is the process of interpretation of information displayed as a family tree. If more than one person in a family is involved with a disease then the pedigree analysis can be done. Specific symbols are used to indicate different aspects of a pedigree as shown in (Fig. 8.5).

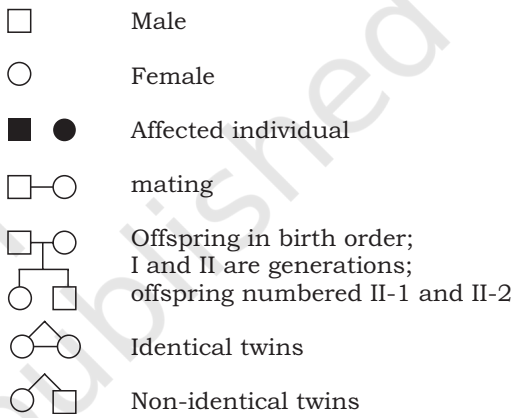


Fig. 8.5: Symbols used in pedigree

Autosomal recessive disorder

The word 'Recessive' indicates that 2 copies of the genes are needed to have the trait and disorder in case of mutated gene. Out of the 2 copies one gene is inherited from father and one from mother. If an individual carries one defective recessive and one normal recessive gene then he or she will be the carrier and not develop the disease. On the basis of statistical projection it is estimated that every human being carries about 5 or more defective recessive genes which can cause a genetic disease. The disease phenotype of a recessive disorder is due to homozygosity of a recessive allele and the unaffected phenotype is secondary to corresponding dominant allele. This can be explained with the example of sickle cell anemia which is a autosomal recessive disease. Sickle cell disease is caused by a mutation in the hemoglobin-β gene found on chromosome 11. This results in a defective haemoglobin

© NCERT not to be republished

(Hb). After giving up oxygen these defective Hb molecules cluster together resulting in formation of rod like structures. The red blood cells become stiff and assume sickle shape (Fig. 8.6).

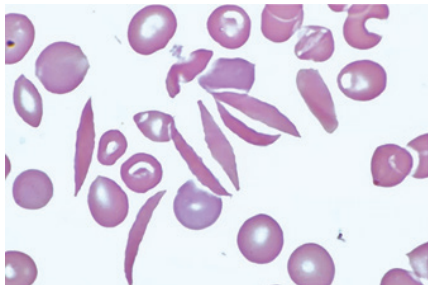


Fig. 8.6: Peripheral blood shows sickling of red blood cells in case of sickle cell disease

Sickle cell anemia is determined by an allele which we can designate as s and the normal condition by S. The person affected by the disease will have a genotype of s/s and unaffected ones will be either S/S or S/s. We can make a projected pedigree of the disease as follows assuming that both parents are carriers (S/s) (Fig. 8.7):

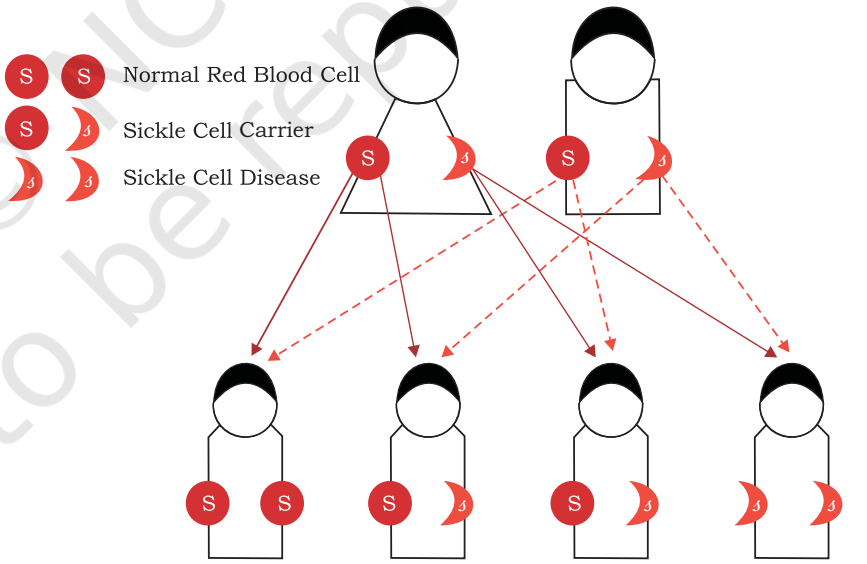


Fig. 8.7: A cross showing inheritance of sickle cell anemia disease

Sickle cell anemia is particularly common among people whose ancestors originated from sub-Saharan Africa, South America, Cuba, Central America, Saudi Arabia, India, and Mediterranean countries. In India it is common

© NCERT not to be republished

among people of the Deccan plateau of central India with a smaller focus in the north of Kerala and Tamil Nadu.

Other examples of autosomal recessive diseases include cystic fibrosis, Tay Sach's disease and phenylketonuria. Individual with cystic fibrosis produce mucus that is abnormally thick and sticky that can damage different organs specially lungs resulting in chronic infections. Tay-Sachs disease is due to absence of an enzyme called hexosaminidase A which results in a fatty substance accumulation in nerve cells particularly affecting the brain. It is a fatal disease manifest in childhood. One in 27 persons of European Ashkenazi Jewish origin individuals carries the Tay-Sachs gene. Phenylketonuria is caused by a mutation in phenylalanine hydroxylase gene resulting in increase in phenylalanine in the blood.

Autosomal dominant disorder

In this kind of inheritance the normal allele is recessive and the abnormal allele is dominant. A rare autosomal dominant disorder Achondroplasia may be considered as an example which leads to a type of Dwarfism in the affected individuals (Fig. 8.8).

In this condition normal people have genotype d/d, affected individual with mild disease have D/d and severly affected have D/D which is often lethal. So most of the surviving cases of achondroplasia are heterozygotes.

Huntington's disease is another example of a rare autosomal dominant disorder which affects nervous system.

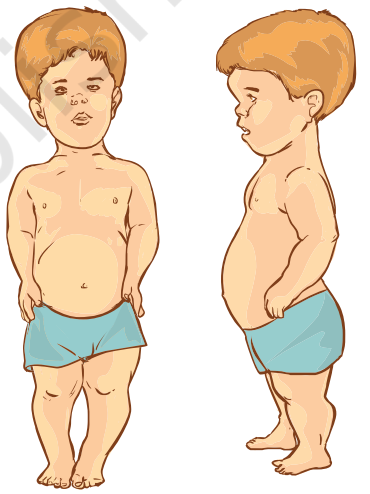


Fig. 8.8 Schematic diagram of Achondroplasia causing Dwarfism

X-linked recessive disorder

In the X-linked recessive inheritance in the mother (XX) the affected gene remains on one X chromosome, as a result she becomes the carrier and usually only males (XY) are affected by the disorder (Fig. 8.9). In the progeny the men pass the Y chromosome on to their sons and their X chromosomes to their daughters. So a man who is affected will not pass it on to his sons but all his daughters will be carriers.

Some examples of X-linked recessive disorders are, Haemophillia, and Duchenne Muscular Dystrophy. Haemophilia is a bleeding disorder associated with mutation in coagulation factor VIII gene (type A) or factor IX

Fig. 8.8 courtesy: https://www.shutterstock.com/image-vector/dwarfism-257159986

© NCERT not to be republished

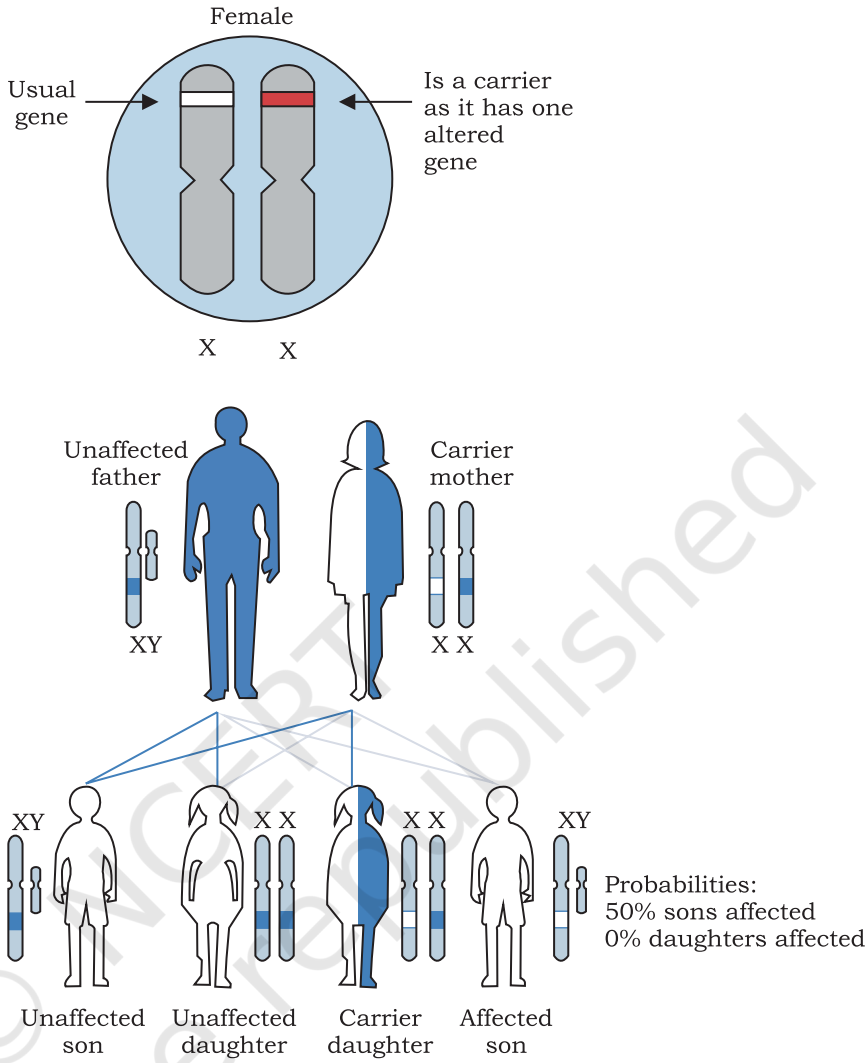


Fig. 8.9: Inheritance of an X-linked disorder

gene (type B). Mutations in coagulation factor genes result in the production of an abnormal version of coagulation factor VIII or IX, or reduced amount of one of these proteins. The altered or missing coagulation factor cannot effectively complete the blood clotting process leading to spontaneous bleeding or increased bleeding tendencies.

Duchenne Muscular Dystrophy (DMD) is caused by mutation of dystrophin gene resulting in reduction/ absence of dystrophin or presence of abnormal protein. Dystrophin abnormality or deficiency leads to dystrophy or degeneration of muscle making them more fragile and weak.

© NCERT not to be republished

X-linked dominant disorder

In this type of inheritance the affected males pass on the mutated dominant gene to all their daughters but to none of their sons (Fig. 8.10a). In case of affected female married to unaffected male the condition is passed on to half of their sons and daughters. (Fig. 8.10b). Examples of such disorders are hypophosphatemia, a type of vitamin D resistant rickets and Alport syndrome, which is associated with progressive hearing loss and kidney disease.

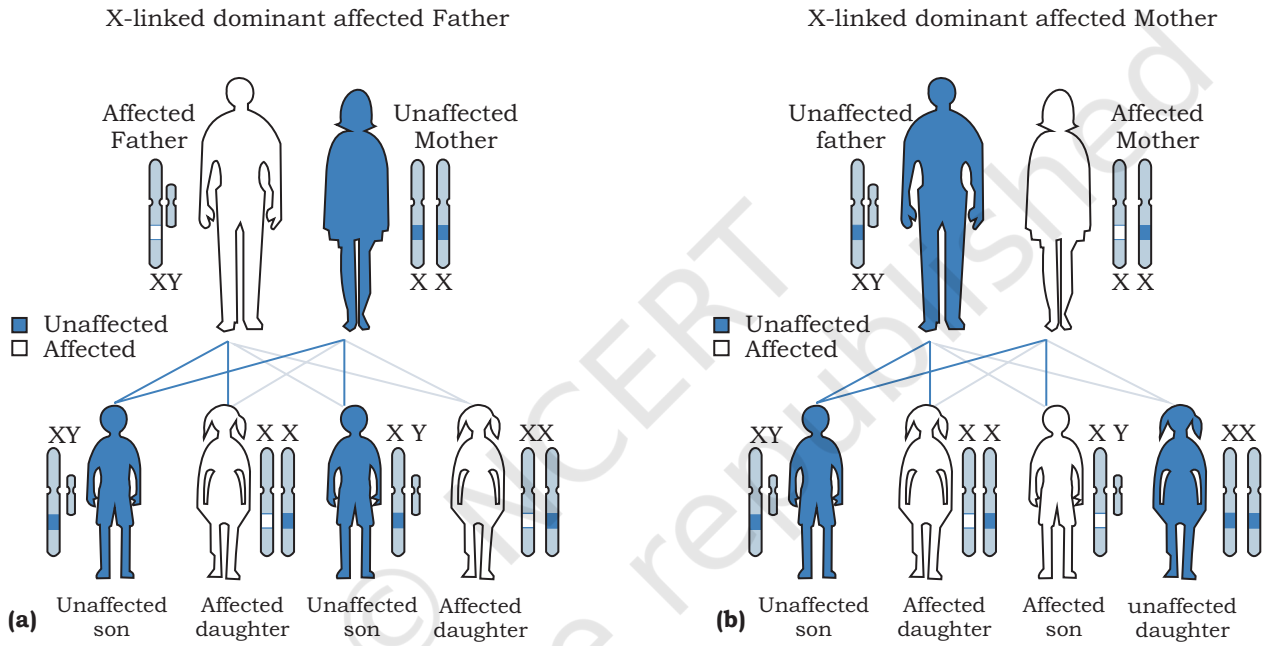


Fig. 8.10 Inheritance of an X-linked dominant disorder through (a) affected father and (b) affected mother

8.3 Polygenic Disorders (Hypertension, Coronary Heart Disease, and Diabetes)

A polygenic disorder is caused by the defect or combined action of more than one gene. The classical examples of polygenic disease include hypertension, coronary heart disease and diabetes. As the pathogenesis proceeds such diseases depend on the simultaneous association of several genes and therefore, they can not be explained as monogenic diseases.

Diabetes mellitus is an important example of polygenic disease. It is a heterogeneous group of disorders

© NCERT not to be republished

characterised by persistent high blood sugar level or hyperglycemia. There are two most common forms of diabetes: Type 1 diabetes (T1D, previously known as insulin dependent diabetes or IDDM) and type 2 diabetes (T2D, previously known as non-insulin-dependent diabetes or NIDDM). Type I diabetes is caused by the destruction of the beta cells of the pancreas due to immunological abnormality. This type constitutes approximately 10% of all cases of diabetes. Life long treatment with insulin injection is needed for these cases. Type 2 diabetes is the commonest form of disease and represents about 90% of cases. It is caused by impaired insulin secretion from beta cells of pancreas and also peripheral insulin resistance. Insulin is needed for transporting glucose from blood to inside the cells like muscle cells for producing energy and maintaining cell function.

Hypertension or presistent increase in blood pressure is a major risk factor for renal, heart and stroke involving brain. This is leading cause of global mortality and morbidity. The normal blood pressure (BP) as <120 (systolic) /<80 (diastolic) mm Hg; elevated BP 120-129/<80 mm Hg; hypertension stage 1 is 130-139 or 80-89 mm Hg, and hypertension stage 2 is ≥140 or ≥90 mm Hg.

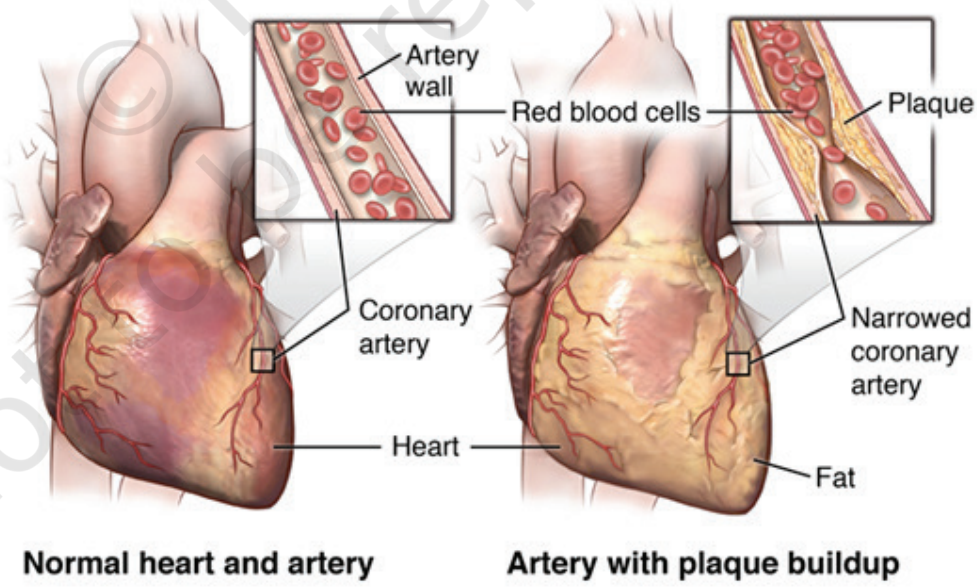


Fig. 8.11: (a) Normal heart and artery and (b) Artery with plaque build-up in coronary heart disease

© NCERT not to be republished

The Coronary Heart Disease (CHD) develops due to narrowing of the coronary artery which supplies blood to heart muscles by gradual build-up of fatty material on its wall (Fig. 8.11). This accumulation of fatty material on the wall of the artery is called **atherosclerosis**. Due to narrowing of the coronary artery by atherosclerosis it can not supply oxygen rich blood to heart muscle which needs it for its continuous activity. This phenomenon of reduction of blood supply is called ischemia. That is why previously coronary heart disease was called as ischaemic heart disease.

Box 2

Mitochondrial inheritance and diseases

As you have studied earlier, mitochondria are organelle present in the cytoplasm of the cells which are primarily responsible for energy production for all cellular activities. Mitochondria produce energy by ATP (adenosine triphosphate). The ATP production depends on a series of well regulated chemical reaction under specific interaction of different enzymes. There are genes in the mitochondria which contain information or codes for these critical enzymes. Any defect or mutation in these genes can affect the ATP production and cellular function. There are variable number of mitochondria in different cell types. Cells in organs like brain, heart, kidneys, muscle and liver which are metabolically very active needs continuous high energy supply. These cells have a large number of mitochondria. In a defect with mitochondrial gene if all mitochondria are affected then it will not be possible for the person to survive. So, in human disease due to mitochondrial gene defect the symptoms and disease severity depends on the ratio of normal and abnormal mitochondria in cells.

Mitochondria are the organelle which contain DNA in circular form, and in animals it is the only organelle in addition to nucleus which contain DNA and gene. The sperm contains very low number of mitochondria and mitochondrial gene. So in the offspring the mitochondrial genes are inherited from the mother. Thus a father with mitochondral gene defect can not transmit the disease to his offspring. The concept of mitochondrial inheritance is explained in the Fig. 8.12.

© NCERT not to be republished

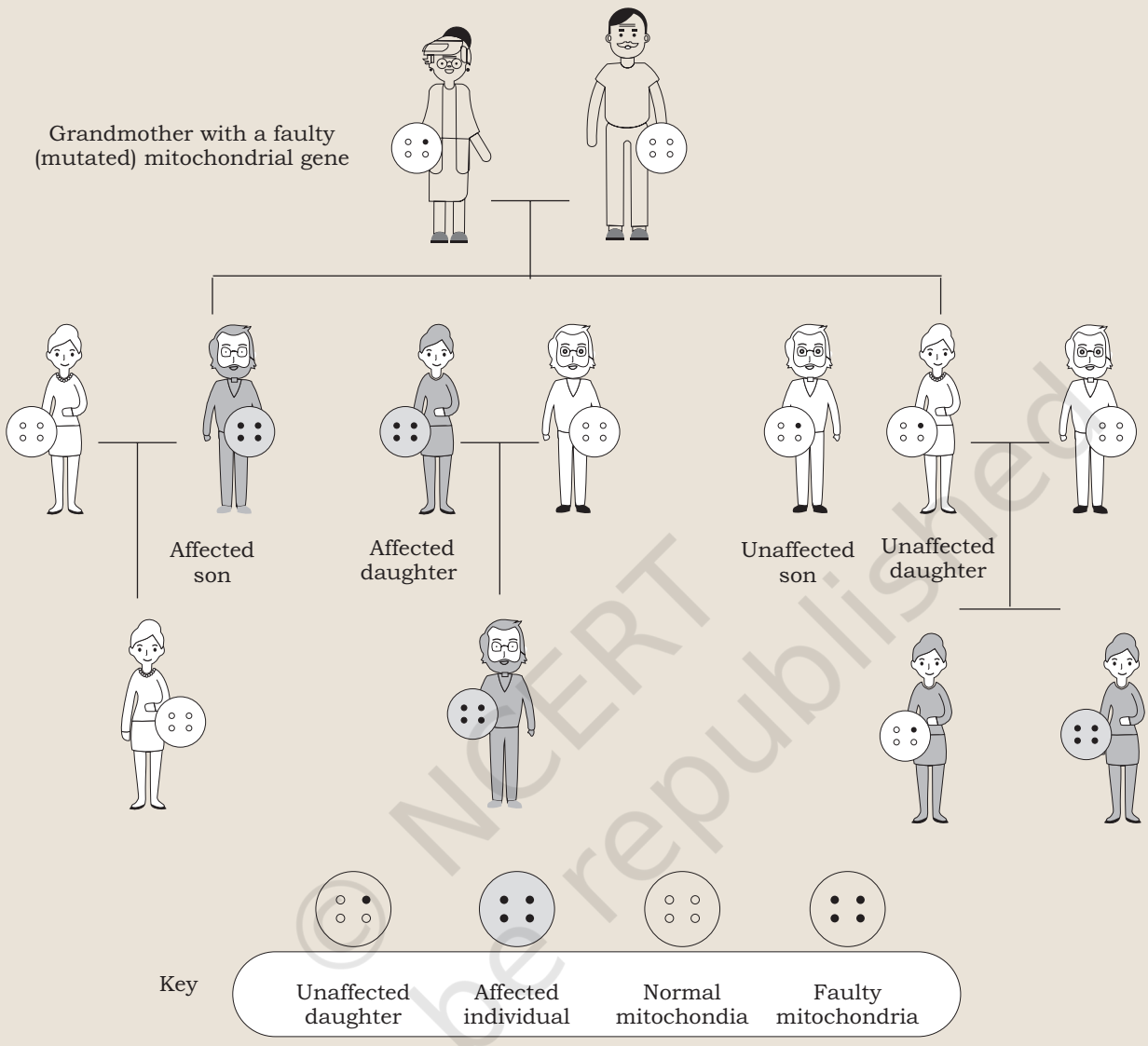


Fig. 8.12: Diagram showing Mitochondrial inheritance in a family with a faulty (mutated) mitochondrial gene

© NCERT not to be republished

SUMMARY

- Haploid cells have only one copy of the chromosome while diploid cells have two copies of the same chromosome.
- Any deviation where one or few chromosomes are either absent or present in multiple copies is called aneuploidy.
- In polyploidy condition, the entire set of chromosome is multiplied.
- A syndrome is a specific collection of signs or symptoms suggesting a particular disease, while a disease is a broader term that refers to abnormal physiological response to internal or external factors.
- Symptoms are subjective and signs are objective.
- Structural chromosomal abnormalities can be caused by deletion, duplication, inversion and translocation.
- Monogenic disease is caused by modifications in a single gene. They can be classified into the following: autosomal recessive, autosomal dominant, X-linked recessive and X-linked dominant.
- Polygenic disorder is caused by the defect or combined action of more than one gene.
- Mitochondrial inheritance and diseases are caused due to defect or mutation in the genes coding for critical enzymes present in the mitochondria.
- Four types of structural chromosomal abnormalities exist: deletion, duplication, inversion and translocation.
- Down syndrome occurs due to trisomy 21 and are characterised by features like moon like face, protruding tongue, muscular hypotonia, palmar crease and so on.
- Klinefelter syndrome occurs due to presence of an extra X chromosome in males and is characterised by tall height, enlarged breasts, coarse voice, hypogonadism and so on.
- Turner syndrome occurs due to absence of one X chromosome in females and is characterised by short stature, webbed neck, small breasts, no menstruation and so on.

© NCERT not to be republished

EXERCISES

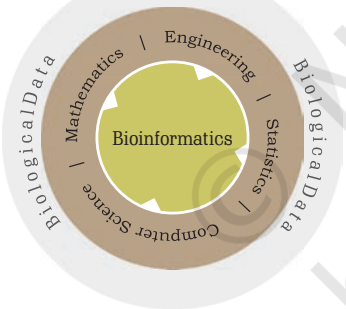
1. Define following terms: dominant, recessive, homozygous, heterozygous, phenotype and genotype.
2. Describe the origin, symptoms and treatment of Down syndrome.
3. Describe the origin, symptoms and treatment of Klinefelter syndrome.
4. Describe the origin, symptoms and treatment of Turner syndromes.
5. Describe various structural chromosomal abnormalities.

© NCERT not to be republished

Chapter 9
Introduction to Bioinformatics

Chapter 10
Protein Informatics and Cheminformatics

Chapter 11
Programming and Systems Biology



Unit IV
Quantitative Biology and Bioinformatics

The recent advances in science have resulted in generation of enormous amount of data obtained from genome sequencing and functional genomics. Handling of such vast data obtained from diverse sources is beyond the scope of humans. This has consequently given rise to a whole new field of science called bioinformatics. Chapter 9 describes the various terminology and concepts of bioinformatics. The scope of utilisation of raw data in biology to retrieve important information about proteins of interest has been discussed in Chapter 10. Chapter 11 deals with the accession, filtration and manipulation of biological data with the help of programming languages.

© NCERT not to be republished

Margaret Oakley Dayhoff (1925-1983)

Margaret Oakley Dayhoff (1925-1983) was an American physical chemist and one of the most important figures in the field of bioinformatics. She received her doctoral degree from Columbia University in the Department of Chemistry and dedicated her entire career to applying mathematics and computational methods to biochemistry. In 1965, she published a comprehensive, open source collection of protein sequences—*Atlas of Protein Sequence and Structure*. It subsequently became a model for the sequence databases which later developed. She had also developed the one-letter codes for amino acids in an attempt to reduce the size of data files in computer applications.

© NCERT not to be republished

11150CH09

CHAPTER 9

Introduction to Bioinformatics

9.1 *The Utility of Basic Mathematical and Statistical Concepts to Understand Biological Systems and Processes*

9.2 *Introduction*

9.3 *Biological Databases*

9.4 *Genome Informatics*

9.5 *Role of Artificial Intelligence (AI) in future*

9.1 The Utility of Basic Mathematical and Statistical Concepts to Understand Biological Systems and Processes

The objective of this chapter is to explain the understanding of the basic concepts of mathematics and statistics is important to a biologist.

The outcome of any biological experiment is data. Previously, biologists used to generate and analyse data without the help of sophisticated software, computational tools and statistical tests. However, this is not the case anymore. With the advent of instruments like high-throughput DNA sequencers, powerful microscopes and other imaging systems, and analytical instruments capable of generating large volumes of data, biologists can no longer deal with the data using their notebooks and excel sheets. Instead, they need computational and statistical tools to handle data. Large volumes of data often require quantitative analyses to interpret and generate biological meaning. Performing such analyses require one to have

© NCERT not to be republished

good working knowledge of computational and statistical concepts, for example; machine learning technologies, regression, variance, and correlation, etc. Mathematical and statistical concepts can only aid biologists to interpret their data and are not a replacement for asking the right questions and the biological acumen. The names of some of the commonly used statistical terms used in biology is provided in Box 1.

Box 1

Box 1: Glossary of the commonly used statistical terms in biology

Null hypothesis— A statement that there is no relationship between two measured phenomena.

Statistical significance— A result has statistical significance when it is very unlikely to have occurred.

p-value— The probability of finding the observed results when the null hypothesis of a study question is true.

t-test—An analysis of two populations means through the use of statistical examination.

Multivariate analysis: A set of techniques used for analysis of data that contain more than one variable.

Regression analysis—A technique to investigate the relationship between a dependent and an independent variable.

Multiple testing correction— A statistical test that corrects for multiple tests to keep the overall error rate to less than or equal to the user-specified P-value cutoff

Analysis of Variance or ANOVA— A collection of statistical models used to analyse the differences among group means in a sample.

Let us examine with specific examples where both the knowledge of computing and statistics can help understand biological phenomena better. For example, we want to understand the association, if any, between blood pressure and heart rates in ten patients (Table 9.1). As provided in the table below, a simple visual estimation (Fig.9.1) is not sufficient to accurately determine the relationship (correlation) between the two variables. For that, one needs to draw a regression line. Correlation and regression are distinct, yet correlated. Correlation quantifies how the variables are connected, but regression defines a statistical relationship between two or more variables where a change in one variable is

© NCERT not to be republished

associated with a change in another. Therefore, in the example above a simple regression test will tell us if there is a direct relationship between heart rate and blood pressure. The output of a linear regression analysis is R^2-value, a statistical measure to show as to how close the data is to the fitted regression line. The R^2 value ranges from 0 (no correlation between the variables) and 1 (perfect correlation between the variables). As shown in Fig. 9.1, the R^2 value suggests that there is a good correlation between the two variables. Therefore, the null hypothesis is rejected in this case.

Table 9.1: Heart rate and blood pressure recorded in ten patients

Patient	Heart rate	Blood pressure (cystolic)
1	112	189
2	83	140
3	92	153
4	121	192
5	85	147
6	111	178
7	94	135
8	88	143
9	102	177
10	111	189

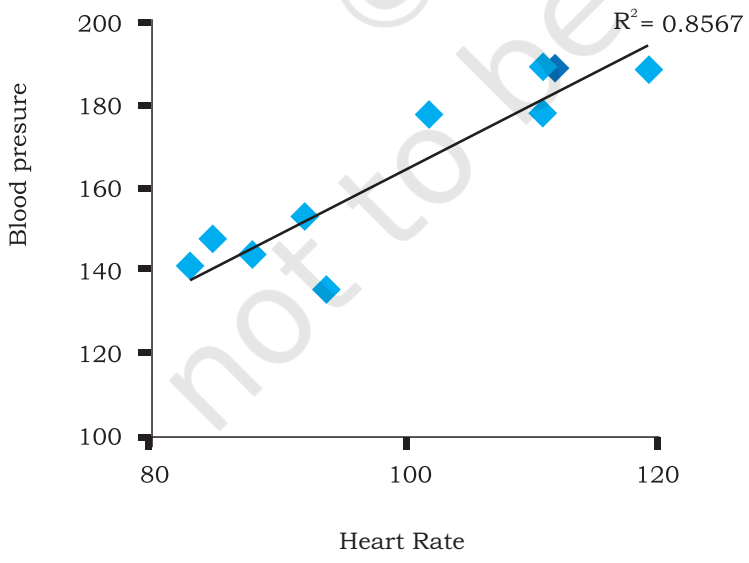


Fig. 9.1: Correlation between the two variables with a simple linear regression line

© NCERT not to be republished

Many fields of biology require a basic understanding of probability. Mathematical modeling of complex systemic phenomena such as cellular mechanisms allows one to understand the vital parameters of the system and its kinetics. Phylogenetic reconstruction, determining ancestral sequences and modeling rates of evolution from a bunch of extant sequences requires knowledge of probability. Biologists need to keep statistical issues before performing an experiment. For instance, choosing an adequate number of samples and replicates, both biological and technical, for the experiment requires knowledge of statistics. An experiment must be repeated multiple times independently to ascertain confidence in the results and to know if they are real or fake. A necessary background in statistical randomness and law of large numbers equips one to deal with this problem. A random sampling from a large number reduces the chance of obtaining biased results. The biologist needs to make sure that the results are statistically significant. This step requires familiarity with various tests and measures of statistical significance and to apply the correct test(s) for the problem in question. Depending upon the problem, the biologist may have to correct and adjust the measure of significance for multiple testing.

For higher levels of computing, analysis and visualisation, a biologist can use the built in frameworks, such as MATLAB (commercial) and R (open source), etc.

For biologists, the choice of the statistical analysis employed is the key to determining the right answer. A weak or incorrect statistical standards lead to false assumptions and therefore may lead to irreproducible results. For example, the commonly used concept in statistics is the P value as the evidence of support for a hypothesis. The smaller the P value, the more likely it is that the result of the test is significant. A P value cutoff of 0.05 (95% significance) or less is considered to be significant. However, the 0.05 threshold has caused too many false positives to appear in the scientific literature. Therefore, the P value cutoff of 0.05 needs to be re-examined. With small sample sizes, one is better off showing all independent data points rather than distorting the visualisation with a misleading average and standard deviation. The statistical

© NCERT not to be republished

power that one begins to consider when faced with a negative result should also be considered in case of positive results. The assumptions of certain established statistical models and distributions to the wrong type of data is, therefore, a common abuse. For example, an assumption of a Gaussian distribution to nonlinear dynamical systems, which results in false positives. Unbalanced mathematical models constructed with unrealistic parameter weights is yet another common abuse and one that is difficult to detect. With due consideration to these caveats, the application of mathematics and statistics to biology can lead to the opening of newer areas of research that are inter-disciplinary in nature to tackle more complex biological problems.

9.2 Introduction

Bioinformatics is an interdisciplinary field that uses computational, mathematical, statistical and occasionally, engineering approaches, in analysing biological information for solving biological problems (Fig. 9.2). Thus, bioinformatics deals with storage, retrieval, analysis and interpretation of biological data using computer based software and tools. Although there are differences, it is alternatingly and interchangeably used with other terms such as 'computational biology,' 'mathematical biology,' 'quantitative biology' and 'bio-statistics,' depending on the dominating disciplinary components. It must be noted, however, that use of these definitions vary among experts and practitioners, and has changed with time.

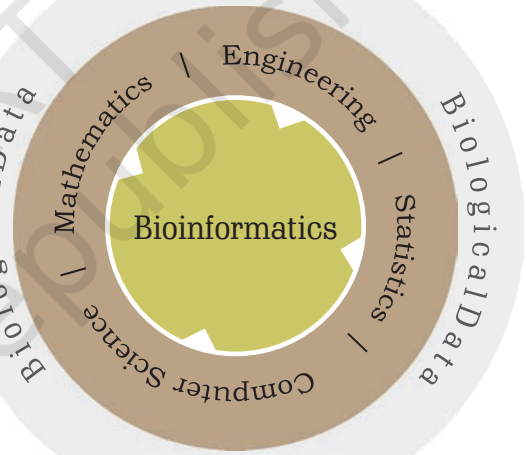


Fig. 9.2: Inter-disciplinary nature of bioinformatics: the intersection of biology with one or more other disciplines such as computer science, mathematics, engineering, and statistics

9.2.1. Historical perspective

Bioinformatics aids in discovery of new findings by data mining as well as the generation of new hypothesis. This is done via modeling or analysis of molecular data. Most bioinformatics tools use either pre-existing nucleotide

© NCERT not to be republished

and protein data from sequence and structure databases, or newly created data generated using high-throughput instruments like next-generation sequencers and DNA microarrays. The National Center for Biotechnology Information (NCBI) in the USA was created as a resource for bioinformatics tools and services. It houses nucleotide, protein, bibliographic databases, etc. GenBank, a widely used database stores all publicly available DNA sequences, was launched in 1982. Although bioinformatics was practiced much before the widespread use of the term, it was not until 1991 when it started appearing in the literature. The name gained wide acceptance after the launch of the human genome project and bioinformatics tools were used extensively for the analysis of sequence data. Therefore, the use of the term bioinformatics in the literature is not more than 30 years old. Bioinformatics has gained broader attraction in the post genome sequencing and high-performance computing era, following advancements and accessibility in biotechnology and computing technology. Before this, when the focus was on lower throughput assays, such as studying the action of a single gene or studying morphology under a microscope, bioinformatics was still used but on a smaller scale.

Structural bioinformatics preceedes informatics based on high-throughput genome-wide assays such as sequencing and DNA microarrays. This is because studies on three-dimensional structures of proteins using NMR spectroscopy and X-ray crystallography in the early 1900s, pre-dates genome and other -ome informatics introduced only in early 2000 and are continuing till date. The number of Protein Data Bank (PDB) structures and GenBank entries are growing every year. The primary concern in bioinformatics is to manage the sequence and structural data in the form of databases, and mine data from these databases to get biological meanings. NCBI hosts nucleotide and protein data, under various categories (Gene, Genome, Structure, Sequence, etc.). Currently, biological data, produced at an unprecedented rate and along with their analysis and interpretation leading to critical biological insights have taken a higher priority. New, optimised and superior algorithms and tools with

© NCERT not to be republished

statistical adaptations and synthesis from multiple fields are developed and implemented to achieve this. Databases based on secondary and tertiary levels of information such as molecular pathways, gene expression, protein structure and function, interaction networks, disease-associated changes, organism specificity and regulatory networks have since been developed and used. Bioinformatics is an evolving field. Due to the dynamic nature of the biological data, genes and exon-intron boundaries and discrepancies in sequences, *in silico* translation errors such as frameshift errors, annotation errors, assembly errors, and simple spelling mistakes are continually being updated.

We will now learn about the different types of biological molecules, kinds of data produced by these techniques, and commonly used analytical and statistical workflows to interpret and visualise data (Fig. 9.3). Details of the experimental techniques used for the production of genomic data is described in Unit V.

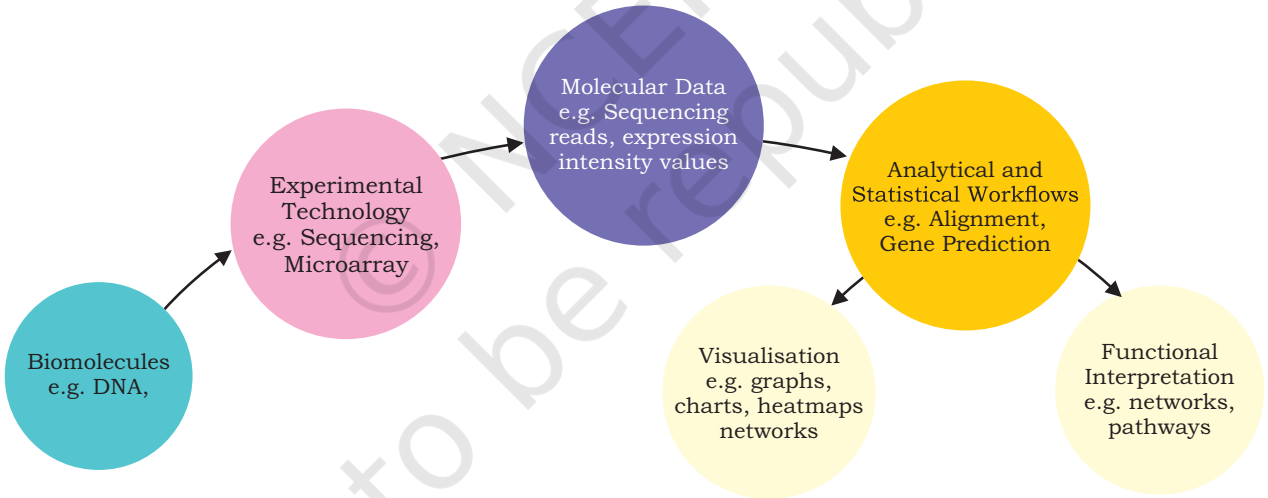


Fig. 9.3. From biomolecules to function

9.2.2. Types of experimental technologies for analysing biomolecules

A few critical experimental technologies used for identification and/or quantification of biomolecules are given in Table 9.2. Details on some of these techniques are described in Unit V.

© NCERT not to be republished

Table 9.2: Name of the technology, biomolecules assayed and the purpose of the technology

Technology	Biomolecule	Purpose
PCR (Polymerase Chain Reaction)	DNA	Amplify a region of interest
RT (Real-Time)-PCR/qPCR (quantitative PCR)	RNA	Detect RNA expression
Next-generation sequencing	DNA/RNA	To sequence genes/genomes and RNA
Gel electrophoresis	DNA, RNA and Proteins	Separation of fragments, based on their size and charge
HPLC (High-performance Liquid Chromatography)	Metabolities	Separation, identification and quantification of metabolites
MS (Mass spectrometry)	DNA, proteins, metabolites, trace gases	Fragmentation, measurement of isotopic composition and mass determination
EM (Electron microscope)	DNA, RNA or protein	Structure, and sequence determination

9.2.3 Types of molecular data

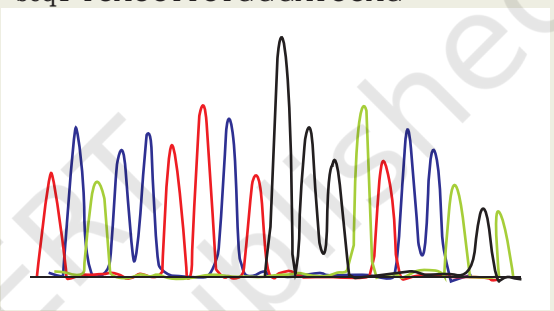
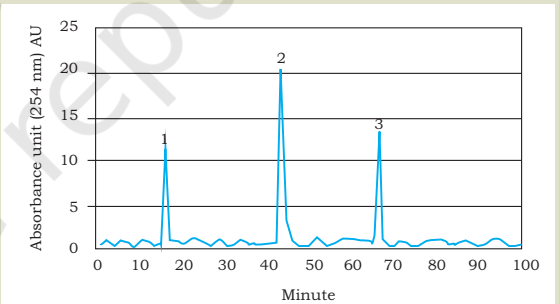
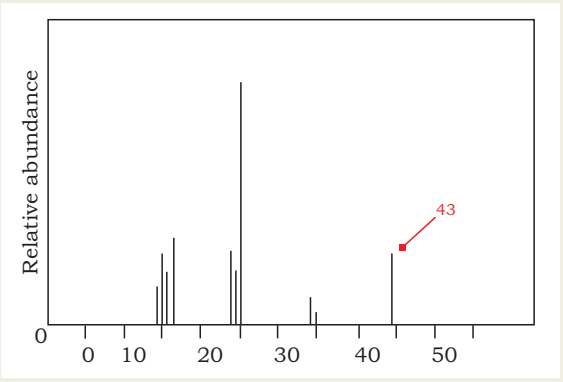
Different technologies assay different biomolecules and produce different types of data output in various formats (Table 9.3). The descriptions on the two commonly used DNA data formats **(FASTQ and FASTA)** are provided later in the chapter (Genome Informatics).

9.2.4 Commonly used analytical and statistical workflows

Biological knowledge may or may not be mandatory for implementation of open source or other proprietary tools. However, it is a must for asking relevant questions and interpretation of biological results as mentioned in the beginning of this chapter. One must understand the logic and principle, and be aware of the underlying assumptions behind the working of any tool.

© NCERT not to be republished

Table 9.3: Types of molecular data and their formats

Technology	Molecular data formats	Examples
Next-Gen Sequencing	Reads (.fastq)	@read1 TTTCCGGGGCCCATAATCTTCTGCAG + 8C;=<=;9@4368>9:67AA<9>63<
Microarray	.cel, intensity data files (.idat), reports binary file	PROBE_ID Assay_Name_1.QT1 Assay_Name_1.QT2 Assay_Name_2.QT1 Assay_Name_2.QT2
Capillary Gel Electrophoresis based Sequencing (Sanger)	.chromat, .fasta, .ab1	>seq1 TCACCTTCTGGGATCCAG 
High performance Liquid chromatography (HPLC)	Chromatogram	
Mass spectrometry (MS)	Mass spectrum	

© NCERT not to be republished

Some commonly used analysis tools are as follows:

- Homology search (Basic Local Alignment Search Tool (BLAST) – blastn, blastp)
- Sequence alignment (CLUSTAL, MAFFT, MUSCLE)
- Phylogenetics (PHYLIP, PAUP)
- Gene prediction (GlimmerHMM, GenScan)
- Functional homology search (HMMER)
- RNA structure (mfold, sFold, uniFold)
- Regulatory region analysis (MatInspector, BEARR, RSAT)
- Protein structure (Phyre2, Jpred)

Bioinformatics tools use various statistical and computational algorithms and approaches. Some commonly used statistical packages are:

- Statistical Package for the Social Sciences (SPSS)
- Statistical Analysis System (SAS)
- R
- Microsoft Excel

At the final stages of the biological data analysis, gene and protein level findings are linked to certain functions. These functional interpretations can be made by using commonly used biological tests like loss- or gain-of function assays, gene knockouts and gene editing. Additionally, inferring affected networks and pathways by using computational tools, one can assign functional significance to a gene and its protein product.

9.3 BIOLOGICAL DATABASES

A biological database is a repository that contains an organised, structured and searchable collection of biological data. In other words, it is a library of biological information, easily accessible and searchable. A biological database links all the relevant data to their original creators or to a reference that describes the underlying data. The information in a database is collected through experiments and computational approaches. For example, a database of human genes contain both the actual nucleotide sequence of all the genes and their characteristics. The database can be made by a single group of researchers collecting information from a variety

© NCERT not to be republished

of public resources or by multiple teams of researchers who can add data to a single repository. A biological database may store multiple types of information, for example, primary nucleotide sequence of a gene; mutation in a given gene specific to a disease and the frequency of single nucleotide polymorphism (SNP) in various population; translated protein sequence of the genes, the 3D structures of proteins and domains, and the functional interaction of one protein with other. The characteristics of a good biological database is one that is easy to access and use, has a user friendly interface, has excellent documentation, has support staff that can answer any queries from its users, lacks errors in the underlying data, cross-referenced, and continually updates information as and when the primary source is updated. There are two main types of databases, relational and non-relational. Databases are managed by using a software system called **database management system (DBMS)** that is used to manipulate, retrieve and manage data. The structured query language (SQL) is the standard application program interface for a relational database. A **non-relational or NoSQL database** does not follow the order of a relational database and is used for large sets of distributed and unstructured data.

9.3.1 What is the need for a biological database?

Imagine when you walk into your school library and ask the librarian about a book. What if the librarian has to physically walk to all the shelves and look for the book? This will take time, and there is no guarantee that the librarian will find the book on the shelf. Instead, if the librarian looks for the book by searching a database that stores all the books using a computer by either using a keyword containing the title of the book or the author or both, the job becomes much simpler. This is why we need databases, to make the search process easy and foolproof. Unlike a library of books where physical items that can be seen with open eyes, for example, books, the nucleotide sequence or genes or protein structures are tiny and can not be searched physically. Therefore, the information must be encoded and stored in a machine-readable format in a database that can be easily searched using an

© NCERT not to be republished

user interface. With the exponential growth of biological data, especially the genome data from various organisms alongside their functions and interactions, it has become imperative to store biological information in databases. Some commonly used biological databases are—

- GenBank (https://www.ncbi.nlm.nih.gov/genbank/): A collection of annotated publicly available DNA sequences.
- PDB (www.wwpdb.org): A collection of 3D structures of proteins, nucleic acids, and complex assemblies.
- UniProt (www.uniprot.org): A collection of protein sequences and function.
- PubMed (https://www.ncbi.nlm.nih.gov/pubmed/): A collection of biomedical literature.
- KEGG (www.kegg.jp): A collection of biological pathways, diseases, drugs, and chemical substances.
- OMIM (www.omim.org): A catalog of human genes and genetic disorders.

In addition to the above, there are organism-specific, disease-specific and secondary databases and are used routinely by biologists.

9.3.2 Data visualisation

Biological data visualisation is an essential aspect of bioinformatics. It involves the application of graphics and data representation and incorporates sequences, genomes, alignments, phylogenies, macromolecular structures, microscopy, and other imaging information. A few examples of data visualisation tools and their use are provided in Table 9.4.

Table 9.4. Data visualisation tools and their uses

Visualisation tool	Use
UCSC Genome browser (https://genome.ucsc.edu/)	An online interactive website to visualise macro- and micro-level genome information on vertebrate and invertebrate species.
KEGG (http://www.genome.jp/kegg/pathway.html) Biocarta (http://www.biocarta.com) Reactome (*https://reactome.org/*)	Visualisation of pathways.
CIRCOS (*circos.ca/*)	Visualisation of data in *circular layout.*

© NCERT not to be republished

EXCEL	Histograms, scatter plots, bubble charts, heat maps
R (*https://www.r-project.org/)*	A software environment for statistical computing and generating graphics.
D3.js (https://d3js.org/)	A JavaScript library for producing dynamic, interactive data visualisations in web browsers.
Phinch (phinch.org/)	An interactive, exploratory framework for visualising biological data.
Integrative Genomics Viewer (IGV, http://software.broadinstitute.org/software/igv/)	A visualisation tool for interactive exploration of large, integrated genomic datasets.

9.4 Genome Informatics

Genome

Genome is the complete set of DNA of an organism, including its genes and the intergenic regions. Genomics is a field of science that deals with the structure, function, evolution, mapping, and modification of genomes. **Genome informatics** is the application of bioinformatics tools to process the outputs of genome-wide assays and technologies, facilitating the interpretation of data and linking them to function. Genomics is one of the omics fields (the other commonly used terms being transcriptomics, proteomics, metabolomics) that has evolved rapidly over the last decade.

Genome information is obtained using high-throughput methods or assays by way of using instruments that provide information on DNA/RNA nucleotide sequence, the variations in genomes, changes in gene expression, profiles of regulatory protein binding to DNA/RNA, and DNA/RNA methylation and various other profile changes. Details on these methods are provided in Unit V. The term high-throughput relates to the process that produces a significant amount of data. The amount of data generated by genome sequencing is substantial. For example, some of the large genome centers in the world like the Broad Institute of the Massachusetts Institute of Technology (MIT) and Harvard produces more than 24TB of genome

© NCERT not to be republished

data per day. Furthermore, the magnitude of such data generated is increasing continously day by day. Although there are not many large genome institutes like the Broad, this gives an idea of the magnitude of genome data produced nowadays. In fact, it is postulated that the computing resources required to handle genome data will exceed those involved in processing Twitter and YouTube data. As the amount of data is significant, genome data need the power of computer science, information technology, quantitative methods and analytics, and statistics for gaining an understanding of underlying complexity, patterns and meaning.

9.4.1 Human Genome Project

An initiative to sequence the complete nucleotide content in human cells started in early 1990s. This initiative is known as the Human Genome Project. The method proposed and used by Fred Sanger was used with modifications for both the publicly-funded initiative headed by the National Human Genome Research Institute (NHGRI) of the National Institutes of Health, USA and the private company Celera Genomics. Both efforts used different methods to sequence DNA. The publicly-funded initiative sequenced human DNA cloned into bacterial-artificial chromosomes and Celera Genomics sequenced randomly cut human DNA, a technique called whole-genome shotgun sequencing method.

The first complete draft human genome was published in 2001. Even the project, which announced the release of the entire human genome in 2003, is still is an unfinished task as the human genome sequence available today has many gaps where the sequence information is not known. Most sequenced part of the genome is in the euchromatic region of the chromosomes and with minimal representations from the heterochromatic regions. The heterochromatic areas mainly consist of repetitive elements, are primarily located in the centromeric and telomeric regions of the chromosomes are difficult to assay with the current sequencing technology. Additionally, due to the repetitive nature of DNA, they are difficult to be assembled to a single location in the chromosome unambiguously.

© NCERT not to be republished

9.4.2 Commonly used data format

One of the challenges of bioinformatics is working with different formats of the resulting data. The bioinformatics community adopts a standard data format of data for the same analytes. For example, most DNA sequencing data (with some exceptions) from the high-throughput DNA sequencers are presented in **FASTQ** format. It is a text-based format that stores sequence information along with its corresponding quality scores. Both the sequence letter and its corresponding quality score are encoded with a single ASCII character. The sequence information in the FASTQ sequence uses the **FASTA** format, which is a text-based format for representing the sequence information in single-letter codes (Fig. 9.4).

An example of a FASTA file and a FASTQ file is given in Table 9.3. The first line in a FASTA file usually starts with a ">" (greater-than) symbol and holds a summary description of the sequence, often a unique library accession number or the name of a gene. A FASTQ file uses typically four lines per sequence. The first line begins with a '@' character, followed by sequence description; line 2 has the raw sequence letters, line 3 begins with a '+' character, and line 4 shows the quality values for the sequence in Line 2. The quality values in line 4 contain the same number of symbols as letters in the sequence presented in line 2. Additionally, both line 1 and line 3 may include optional sequence identifiers. In line 4 representing quality values, the character '!' and '~' represent the lowest and the highest quality respectively.

(B) FASTA format

Accession number | Identifier (Gene name) | Source | Type of sequence

>K02382.1 Human growth hormone synthetic gene, complete cds
CGATATGTTCCCAACTATTCCACTGAGTCGCCTGTTCGATAACGCGATGCTGCGTGCGCATCGTCT
GCACCAACTGGCTTTCGACACTTACCAGGAGTTCGAAGAAGCATACATCCCGAAAGAACAGAAATA
CAGCTTCCTTCAGAACCCACAGACCTCGTTGTGTTTCTCTGAAAGTATCCCGACCCCTTCTAACCG
CGAAGAGACCCAGCAGAAATCGAACCTTGAACTGCTTCGTATCTCGCTGCTTCTCATTCAGTCGTG
GCTGGAGCCAGTACAGTTCCTGCGTTCGGTTTTCGCAAACTCACTGGTTTACGGTGCGTCTGACAG
TAACGTTTACGACCTGCTGAAAGACCTTGAAGAAGGGATCCAGACCCTGATGGGTCGCCTGGAAGA
TGGTTCACCACGCACTGGTCAGATCTTCAAACAGACTTACTCCAAATTCGATACTAACTCTCATAA
CGATGATGCTCTGCTGAAAAACTACGGCCTGCTGTACTGTTTCCGTAAAGATATGGATAAAGTTGA

Fig. 9.4: FASTA format

© NCERT not to be republished

9.4.3 Genome Informatic Tools

Genome informatics tools came up in parallel with the development of sequencing technology to cater the analysis of resulting data. The high-throughput sequencing instruments produce sequence reads, either short (about 100-150 nucleotides) or long (few kilobases) depending on the target size and device used. The resulting sequenced reads need to be either assembled into a genome (where no prior genome information is known) or aligned to a reference genome (in the case of re-sequencing). Broadly there are two possible analytical workflows, one based on aligning the reads to a reference sequence (e.g., genome) and second based on the *de novo* assembly of the reads into a draft reference genome sequence. In both the scenarios, the sequencing data is pre-processed and checked for quality (Table 9.5).

Table 9.5: Quality Control tools for pre-processing of raw sequencing data

Tool category	Examples of tools	Function
QC (Quality control)	FastQC	To QC high-throughput sequence data
	Trimmomatic	Quality and adaptor trimming

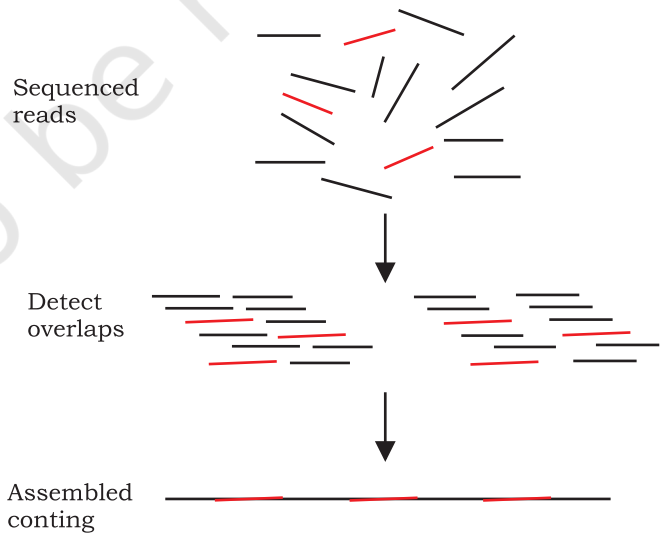


Fig. 9.5: Assembly of reads into a continuous stretch (contig)

© NCERT not to be republished

The *de novo* assembly based workflow involves assembling a genome by piecing together reads based on the overlaps and insert size (distance between paired-end reads). The assembly undergoes further annotation and processing, i.e., prediction of novel genes, and identification of genes based on homology to the known ones, quantification of gene expression, identification of splice variants, novel isoforms, and fused transcripts.

A simplified demonstration of the assembly of reads into a contiguous stretch (contig) is provided in Fig. 9.5.

There are several conversion and auxiliary tools for downstream analysis of genome data. Table 9.6 below highlights a few of those tools.

Table 9.6: Conversion and auxiliary tools used in the analytical workflow

Aligners	Database search	Basic Local Alignment Search Tool (BLAST) -blastn, blastp	Seach nucleotide or protein query sequence against non-redundant sequence database
		BLAST-like alignment tool (BLAT)	Stores the genome index in memory designed to quickly find highly similar sequences (>95%), of at least 40 nucleotides
	Pairwise alignment	BLAST (bl2seq)	Allows comparison of one sequence against the other.
		CLUSTAL	Series of multiple sequence alignment programs
		MUltiple Sequence Comparison by Log-Expectation (MUSCLE)	Better speed and accuracy than CLUSTAL and T-Coffee
		Tree-based Consistency Objective Function For alignment Evaluation (T-Coffee)	Progressive approach supports the use of PDB files in its advanced versions can identify motifs
Conversion tools	SAM tools		SAM (Sequence Alignment/Mapping) to BAM (Binary Alignment/Mapping) conversion
	PICARD tools		SAM to Fastq

© NCERT not to be republished

9.4.4 Downstream analyses, interpreting data and linking to function

Genome annotation

Genome annotation is the process of identification and classification of all the features in a genome. After a genome is assembled, it is analysed for predicting and identifying locations of coding parts of genes, exon-intron junctions, repeat elements, non-coding elements, and pseudogenes. Repeats are a significant component of any eukaryotic genome. They provide diversity to individual genomes and could play a significant role in their evolution. Repeats in a genome are classified into three major categories — transposons, satellites and low complexity regions. Out of these, transposons form the most significant chunk of repeat elements.

Gene prediction

Gene prediction involves identification of all the coding sequence in a genome. Tools used for gene prediction may follow either *ab initio*, similarity-based or integrated model. The similarity-based approach uses sequence similarity for identifying coding regions. Programs like BLAST and its sub-packages are used to carry out an exhaustive comparative search using the existing sequences. The last approach to identify genes utilise both strategies, and integrate the gene prediction results from both of these methods.

Gene ontology

After gene prediction, the gene products, and their associated functions need to be integrated and represented without ambiguity. The Gene Ontology aims to describe gene products regarding their associated biological processes, cellular components, and molecular functions.

Transcript prediction

Prediction of transcripts (both coding and non-coding) are important to understand the identity of unannotated genes/transcripts. There are several tools that can do this. For example, Cufflinks, cuffmerge, and cuffcompare are a part of a suite that helps in the identification of transcripts.

© NCERT not to be republished

Phylogenetic analysis

Phylogenetic analysis is done to determine the evolutionary relationship (phylogeny) of an organism with others. This relationship can be expressed as a diagram called **cladogram** or **phylogenetic tree**, and the distance of an organism on the diagram indicates how close that organism is related to others on an evolutionary scale. There are different methods for calculating evolutionary distances which are used by tools for phylogenetic analysis like PHYLIP and PAUP.

9.5 Role of Artificial Intelligence (AI) In Future

We are moving into an era where the prospect of Artificial Intelligence (AI) is beginning to take hold. Bots such as Libratus and AlphaGo have already conquered real-life experts in games such as Poker and GO. Machine learning is also starting to make its presence felt in improving farming and broadening healthcare options. For example, AI-based tools are used in the USA to diagnose eye diseases that rely on thousands of stored images of the eye. It is possible that we will have computers read X-rays and pathology slides routinely in the hospitals and aid radiologists and pathologists in correctly and quickly diagnose human diseases. AI also holds significant promise to help our farmers to improve their yield and decide on their crops. Many existing bioinformatics tools are already using machine learning algorithms for calling genomic variants and assessing their significance levels. However, similar to other real-life situations where Siri (from Apple) or Alexa (from Amazon) as personal assistants are not perfect and cannot do everything, a successful intersection between AI and bioinformatics will require time and multi-disciplinary efforts to yield successful results. Such a junction demands to build sufficient intelligence in the tools to interpret data and generate hypotheses. The competency of the tools being developed for analysis is still not at par with the fast pace at which the data is being produced. A collaboration among researchers working in biology, computer science, bioinformatics, statistics and artificial intelligence researchers will lead to successful tools for biological data analyses in the future.

© NCERT not to be republished

SUMMARY

- Bioinformatics is an interdisciplinary field that utilizes computational and statistical techniques to analyze and interpret biological data. It involves storing, retrieving, and analyzing biological information, often using computer-based tools.
- Modern biologists require advanced computational and statistical tools to manage, analyse and interpret large datasets, using key statistical methods such as correlation and regression thereby playing essential roles in deriving meaningful insights from experimental data.
- Historical perspective traces bioinformatics from early structural bioinformatics, focusing on protein structures, to contemporary genome- wide assays driven by sequencing technology advancements. Milestones include the creation of GenBank in 1982 and the launch of the Human Genome Project in 1991, which significantly advanced bioinformatics.
- Various biological databases, such as nucleotide and protein databases, facilitate organizing vast biological data. One of the major database namely, National Center for Biotechnology Information (NCBI) is key for representing and assessing sequence data in common data format such as FASTA.
- Genome informatics tools and workflows are vital for analyzing sequence data generated by high-throughput sequencing technologies.
- The potential applications of artificial intelligence (Al) in bioinformatics are discussed in sports, healthcare and agriculture. Al tools are increasingly employed for tasks like diagnosing diseases and improving crop yields, with ongoing multidisciplinary collaboration essential for developing intelligent bioinformatics tools.

© NCERT not to be republished

EXERCISES

1. Name the two modalities of analysis following sequencing.
2. Name any three major types of variants.
3. What are disease-specific variants termed?
   (a) somatic
   (b) germline
4. Which is the preferred tool for transcriptome assembly, in the de novo and genome-guided modalities?
   (a) Tophat2
   (b) Trinity
5. What is the difference between BLAT and BLAST?
6. What came first? Structural Bioinformatics or Genome informatics?
7. Name any two of the major classes of biological macromolecules.
8. DNA sequences can be represented by which of the following data format?
   (a) FASTQ
   (b) FASTA
   (c) AB1
   (d) All of the above
9. Can a phylogeny be produced directly from a multi-fasta file? Justify your answer.
10. Which tool can help you visualise variants in a circular manner?
    (a) UCSC Genome Browser
    (b) CIRCOS
    (c) IGV
11. Why do we need to sequence nucleic acids? What can one gain by understanding the sequence of nucleic acide?

© NCERT not to be republished

11150CH10

Chapter 10

Protein Informatics and Cheminformatics

10.1 Protein informatics
10.2 Cheminformatics

10.1 Protein Informatics

10.1.1 Introduction

Collecting information about any protein using techniques of information technology comes under protein informatics. Protein informatics has been of tremendous help in getting the geometrical location of the functional site, the biochemical function and the biological function of the hypothetical proteins. In addition, it has led to the determination of the tertiary structures of many hypothetical proteins, whose molecular functions could not be understood using conventional methods. Heterogeneous databases and various descriptors of amino acid sequences, tertiary structures and pathways on the proteome scale have also been of help in developing protein informatics.

© NCERT not to be republished

10.1.2 Protein data types

The process of computation of information extraction needs raw data of protein. These protein data can be of following types —

(i) Microscopic image of heat-denatured protein aggregate
(ii) Protein in solution form
(iii) Protein sequence as output of Matrix Assisted Laser Desorption Ionisation (MALDI)
(iv) Assembled protein sequence
(v) Protein crystal structure in Protein Data Bank (PDB) format
(vi) Protein-protein, protein-ligand or protein-nucleotide interaction file
(vii) Nuclear Magnetic Resonance (NMR) data and Mass Spectrometry (MS) data
(viii) Protein sequences derived directly from the genomic sequences, which do not contain the known evidence of existence (Hypothetical protein)

The above mentioned types of protein data can be used for getting useful information like

(i) Multi-fractal property of microscopic image of heat-denatured protein aggregate is used for designing protein-marker.
(ii) Protein data in solution are useful for analysing physico-chemical properties and kinetics information.
(iii) Fragmented short sequences of proteins from MALDI are used to find out the full length sequence.
(iv) Protein crystal structures are used to study mutations and interactions.
(v) PDB, NMR and MS data are also used for the prediction of structure of non-crystallised protein (directly from the sequence).
(vi) There are proteins which do not have known existences (hypothetical proteins) which can be identified from the genomic sequences.
(vii) Network mapping of protein provides information about the possible target of treatment of different diseases.

© NCERT not to be republished

In order to carry out the protein informatics analysis, the following two basic facilities are required:

(i) Availability of the raw data from various databases, such as NCBI, PDB, CHEMBL, BIOMODELS, etc.

(ii) Informatics tools and techniques used for the analyses. Some of the well known techniques are: (a) image analysis by the wavelet techniques, (b) sequence similarity and homology calculations, (c) structure optimisation techniques, (d) data analysis by statistical and machine learning techniques as Artificial Neural Network (ANN), Support Vector Machine (SVM) and Hidden Markov Model (HMM), (e) Network Mapping Technique, and (f) Systems Biology Mark-up Language (SBML).

10.1.3 Computational prediction of protein structures

Protein structure prediction using bioinformatics tools is aimed to explore how amino acid sequences specify the structure of proteins and how these proteins bind to substrates and other molecules to perform their functions. This task for predicting structure of a protein (including those of hypothetical proteins) using bioinformatics tools is possible even when only gene sequence is known, i.e., in the absence of protein sequence. Many computational tools are available from different sources for making predictions of structural and physico-chemical properties of proteins. The major advantages of computational methods are the time frame involved, low cost and the feasibility of high throughput screening.

10.1.3.1 Primary structure prediction

Protein primary structure prediction involves physico-chemical characterisation such as isoelectric point, extinction co-efficient, instability index, aliphatic index and grand average hydropathy. All these can be calculated with the help of ProtParam tool of ExPASy Proteomics Server. Some of the physico-chemical properties of proteins are described in brief in the following section.

Isoelectric point— Isoelectric point (pI) is the pH at which the surface of protein is covered with charge but net charge of protein is zero. At pI, proteins are stable and compact. If the computed pI value is less than 7 (pI<7), it indicates that protein is considered as acidic.

© NCERT not to be republished

The pI greater than 7 (pI>7) reveals that protein is basic in character. The computed isoelectric point (pI) will be useful for developing the buffer system for purification by isoelectric focusing method.

The aliphatic index— The aliphatic index (AI), which is defined as the relative volume of a protein occupied by aliphatic side chains (A, V, I and L) is regarded as a positive factor for the increase of thermal stability of globular proteins. Very high aliphatic index of protein sequences indicates that protein may be stable for a wide temperature range.

The instability index—The instability index provides an estimate of the stability of protein in a test tube. There are certain dipeptides, the occurrence of which is significantly different in the unstable proteins compared with those in the stable ones. This method assigns a weight value of instability. Using these weight values it is possible to compute an instability index. A protein whose instability index is smaller than 40, is predicted as stable, a value above 40 predicts that the protein may be unstable.

The Grand Average Hydropathy (GRAVY) value— The Grand Average Hydropathy (GRAVY) value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. The low range of GRAVY value indicates the possibility of better interaction with water.

10.1.3.2 Secondary Structure Prediction

The protein secondary structure has been studied intensely, since it is very helpful to reveal the functions of protein with unknown structures. In addition, it has been shown that the prediction of protein secondary structure is a step towards protein 3-dimensional structure prediction. APSSP, CFSSP, SOPMA, and GOR are common protein secondary structure prediction tools.

10.1.3.3 Three dimensional (3D) Structure Prediction

The following three computational methods are commonly used to predict protein 3D structure.

Homology modelling—For homology modelling, the amino acid sequence of a protein with unknown structure is aligned against sequences of proteins

© NCERT not to be republished

with known structures. High degrees of homology (very similar sequences across and between the proteins) can be used to determine the global structure of the protein with unknown structure and place it into a certain fold category. Lower degrees of homology may still be used to determine the local structures, an example being the Chou-Fasman method for predicting secondary structure. An advantage of homology modelling methods is lack of dependence on the knowledge of physical determinants. MODELLER and SWISS-MODEL are commonly used tools for homology modelling.

Fold prediction—With the method called 'threading', the sequence of a protein with unknown structure is forced to take the conformation of the backbone (protein side chains) of a protein with known structure. These methods tend to be more compute-intensive than homology modelling methods, but they give more confidence in the physical viability of the results. **LIBELLULA** and **Threader** are commonly used tools for this method.

***De novo* protein structure prediction:** It is an algorithmic process by which protein tertiary structure is predicted from its amino acid primary sequence. **QUARK** is a computer algorithm for *ab initio* protein structure prediction and protein peptide folding, which aims to construct the correct protein 3D model from amino acid sequence only.

Computationally elucidated structure of a protein is recorded as atomic coordinates in protein-data-bank files. The three-dimensional coordinates are stored in a type of text-file namely PDB-file with file extension .pdb in **Protein Data Bank** (PDB) database. It contains data from X-ray crystallography, NMR and a few theoretical structure models.

Domain prediction— Domain is distinct functional and/or structural units of a protein. Independent folding unit of a polypeptide chain also carries specific function. They are often identified as recurring (sequence or structure) units, which may exist in various contexts. Domains provide most valuable information for the prediction of protein structure, function, evolution, and design. The most common tools for domain prediction are **InterPRO** scan of **EMBL** and **CDD** search of **NCBI**.

© NCERT not to be republished

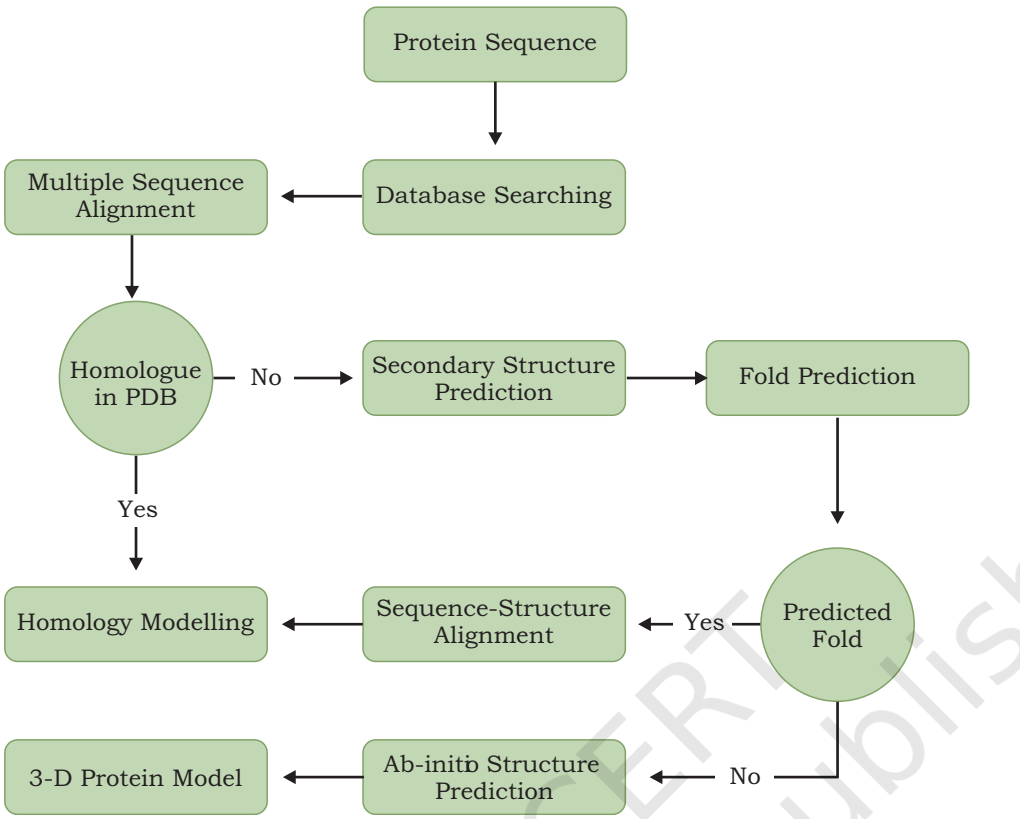


Fig. 10.1: Flowchart of all possible ways for protein structure prediction from a protein sequence

A flowchart depicting various possible ways for protein structure prediction from a protein sequence is shown in Fig. 10.1.

10.2 Cheminformatics

10.2.1 Introduction

The use of computational and informational techniques to understand problems of chemistry is known as cheminformatics. This is an interface science for combining principles of physics, chemistry, biology, mathematics, biochemistry, statistics and informatics. Terms like chemoinformatics and chemical informatics are used along with cheminformatics, to indicate the same approach. Cheminformatics strategies are useful in drug discovery where large number of compounds are evaluated for interaction with the target cellular molecules.

© NCERT not to be republished

For the last two decades, the science of cheminformatics has grown conceptually and technically, finding widespread applications in chemical industry, pharmaceutical and biotechnology research e.g., computer-aided drug design (CADD) where one looks for molecules with specific biological and therapeutic properties.

Cheminformatics specialists handle information on physical properties, three-dimensional molecular and crystal structures, chemical reaction pathways and so on. In addition to real compounds, cheminformatics researchers primarily handle virtual libraries of chemical databases that can contain hypothetical compounds. Virtual libraries can contain information on likely synthesis methods and predicted stability of the reaction products. Virtual screening uses chemical and physical principles to identify and evaluate the best candidates for a particular property or reaction from large libraries of real and virtual molecules. The most desirable candidates can then be verified in laboratory studies.

10.2.2 Storing and managing the chemical data

Many groups and organisations maintain database of chemical compounds, some of them are publicly available for free and some of them are commercially available. Though these databases contain millions of chemical compounds, their reactions and so on, the computational power and tools are so robust that it takes only a few seconds to search through the entire resource and retrieve the records.

Science has advanced so much that we have library of virtual molecules (runs into billions of entries) —these are compounds that do not exist according to the available literature, but can be synthesized using advanced combinatorial techniques.

CAS **(Chemical Abstracts Service)**, a division of American Chemical Society) is the world's largest collection of chemistry insights. It is an authoritative source of chemical names, structures and serves as a universal standard for chemists.

At present, CAS registry hosts more than 219 million organic and inorganic substances taken from literature. The registry includes more than 70 million protein and nucleic acid sequences. The database contains more than 8 billion property values of substances.

© NCERT not to be republished

Data from large number of globally published literature including biomedical sciences, chemistry, engineering, material science and so on, are added to the CAS database everyday. This tremendous resource is a treasure for finding compounds of therapeutic and industrial importance. Some of the popular chemical databases are mentioned in Table 10.1.

Table 10.1: Popular Chemical Database

Name	Description
PubChem	PubChem is a database of chemical molecules which maintains three types of information namely, substance, compound and BioAssays.
ZINC	ZINC database contains large number of compounds available for virtual screening. In this database, various molecule features like molecular weight, log P etc. are included.
ChEMBL	This database provides comprehensive information a large number of bioactive (small drug-like molecules) compounds with drug targets.
NCI	NCI database had large number of small molecule structures, a very useful resource for researchers working in the field of cancer/AIDS.
ChemDB	It is a databse of chemicals containing information of chemicals, which includes predicted or experimentally determined physicochemical properties, such as 3D structure, melting temperature, and solubility.
ChemSpider	ChemSpider contains unique chemical entities aggregated from diverse data sources.
BindingDB	It is a binding affinity database of small molecules for their protein targets.
DrugBank	The database that combines detailed drug (i.e., chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e., sequence, structure, and pathway) information.
PharmaGKB	It is a pharmacogenomics knowledge resource that encompasses clinical information of drug molecules.
SuperDrug	This database contains 3D-structures of active ingredients of essential marketed drugs.

10.2.3 Why do we need cheminformatics?

Faced with hundreds of millions of compounds, properties, chemical reactions and so on, the question is how to navigate this huge resource and find the right chemical

© NCERT not to be republished

compound that meets our requirements?

The cheminformatics tools help us browse through the enormous body of literature and find patterns. Pharma companies use cheminformatics resources and tools for *in silico* design of novel drugs, followed by synthesis and testing. Chemical manufacturing industry needs cheminformatics to design new properties, predict efficacy and toxicity of chemicals before they reach the market.

10.2.4 How to store information on chemical compounds?

One can easily draw chemical compounds on the paper with bonds in between the atoms and aligned at a certain angle. Using drawing tools, it is possible to use predefined templates on the user interface and draw standard geometric structures and reactions with ease. One can store such information as an image file (e.g, jpg, tif) or document form (e.g., doc, pdf). However, such a storage of chemical data is of little use in research projects that demand 'deep browsing' into bond angles, flexibilty of rotation and so on, to find the right molecule for a particular purpose.

Chemical structures are therefore stored in the computer as molecular graphs. A graph is an imaginary representation of nodes (units of chemical substances) and edges (movement of information between nodes). Using node-edge approach, one can create graphs representing atoms and bonds. At a higher level, the same representation is used to build molecular pathways in the cell e.g. glycolysis and Kreb's cycle, etc.

10.2.5 Searching the structures

It is common for commercially available databases to have their origins in academic research projects. This is true for Cheminformatics as well.

The simplest task is to extract information about chemical structure. For example, if you're interested in discovering the physical and chemical properties of a substance, it can provide you with information on all the chemical substances that fall within a specific boiling point range and more.

Another aspect of the search process is substructure retrieval. For example, it displays the various chemical

© NCERT not to be republished

compounds associated with specific functional groups, such as a methyl group, benzene ring, or alkene backbone. We refer to the complete embedding of a small graph into a larger one as subgraph isomorphism (iso, meaning multiple forms of the same type).

For this reason, people often perform a two-stage search. The first stage/step involves filtering out molecules that may not be a potential match for the substructure query. During this stage, we eliminate most of the molecules, leaving behind a small fraction that may be worth further investigating in the next stage. The second stage or step involves a more elaborate subgraph isomerism process to find molecules that truly match a given substructure. We implement molecule screens using binary strings of 0s and 1s, known as bitstrings.

10.2.6 Searching the reactions

While planning a synthesis, a chemist may wish to search the reaction database for products to find if someone has already synthesised a given compound and, if yes, what were the reaction conditions? Also, one may want to know how many different reaction pathways exist to move from point A to point X in the pathway. Further, one may need information on solvents, pH, temperature, pressure and so on. One can refine reaction queries by integrating several queries into one statement: find all reactions that utilise glucose and operate within temperature range of 37° C.

A key feature of reaction search is atom mapping, i.e., finding an exact correspondence between reactant atoms and the resultant products. The existing cheminformatics tools and databases also allow retrieval of reactions where a certain substructure is converted into products.

10.2.7 Pharmacophore

A pharmacophore is a description of molecular features that define molecular recognition of a ligand. The IUPAC defines pharmacophore as an ensemble of steric and electronic features necessary to ensure optimal interactions with specific biological target and trigger a biological response.

A pharmacophore model explains how structurally diverse ligands can dock on to a single receptor molecule. A 3D pharmacophore is a set of features related to spatial orientation e.g., positively and negatively charged groups,

© NCERT not to be republished

rings and hydrophobic regions.

It is important to know that a pharmacophore is not a physical molecule or a group of molecules. The pharmacophore is a well established conceptual framework that defines specific molecular description (pharmacophore points e.g., steric, electrostatic and hydrophobic properties) of a therapeutic molecule needed for its interaction with a target.

10.2.8 Lipinski's rule of five (R05)

This rule was proposed by Christopher A. Lipinski in 1997 and describes the key molecular properties of compounds.

Ideally a drug should be biodegradable, non-toxic, stable, with no side effects, have uniform cellular distribution, controllable release in the body, cost-effective and easily excreted after action.

The Lipinski's rule of five includes the following criteria for finding an orally active drug and it should not bear more than one violation.

(i) Not more than 5 hydrogen bond donors
(ii) Not more than 10 hydrogen bond acceptors
(iii) Molecular weight below 500 Daltons
(iv) Octanol water partition coefficient log P of less than 5

It is important to remember that Lipinski's rule of five only deals with finding a chemical compound that has potential to be a successful oral drug. The RO5 may not apply to the drugs given through intramuscular and intravenous routes.

Based on the RO5, the predicted chemical compound is assigned a value between 0-4. The chemical compound having value less than 3 is not suitable for further analysis. The RO5 does not apply to natural products and semi-synthetic natural products.

10.2.9 The journey of a drug

Nature provides an immense store house of active compounds with therapeutic applications. Using scientific methods, we have learnt how to narrow down to a certain set of compounds that might make promising molecules that one is looking for. The road to drug discovery and development is long, expensive and risky. Fig. 10.2 depicts the overall drug discovery pipeline, i.e., from lab to the market. Virtual screening is an *in-silico* approach to decide

© NCERT not to be republished

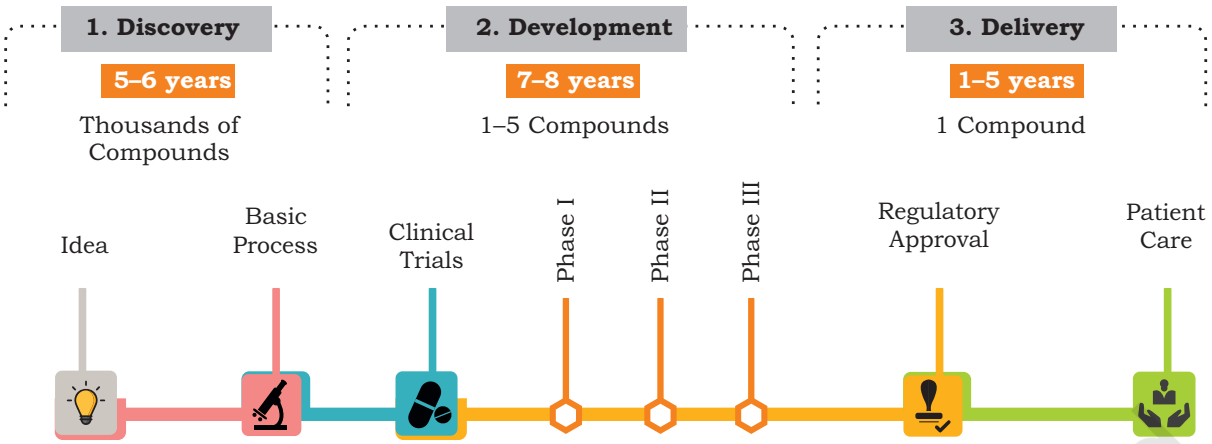


Fig. 10.2: Drug development pipeline from lab to market

Box 1

1. In the early 1990s, Pfizer tested out a drug called UK92480, which is expected to relax the blood vessels and treat heart complications. Instead, they found an unexpected side effect on reproductive system and developed the drug into a blue pill known as viagra.
2. Do you know the origin of saccharin, an artificial sweetner that we often use in the tea or coffee? The origin of this discovery was accidental. One day in 1879, a Russian Chemist Dr. Constantin Fahlberg, was so involved with his work that he forgot about the supper till it was quite late and rushed off for a meal without washing the hands. He broke a piece of bread with unwashed hands and put it in the mouth. It tasted very sweet. At this time, he did not suspect much. Rather, he washed his mouth and dried moustache with a napkin. Interestingly, napkin also tasted sweet. Then he drank water and that also tasted sweet. Immediately, he sensed the breakthrough of sweetner coming from coal tar. He dropped the dinner, ran back to the lab and tasted the contents of every beaker. One of them had an impure solution of saccharin. Luckily there wasn't any corrosive liquid. He worked on this for months, found a chemical method of making saccharin, started a company and became rich and globally famous for this breakthrough.

which compound among billions is useful for a particular purpose. The purpose may be related to drug discovery, industrial applications and so on.

In the virtual screening, one scores, ranks and extracts a set of structures using computational methods. The virtual screening may consist of series of filters that eliminate undesirable compounds at every step. As we move from the first step of virtual screening to the last step, the criteria gets increasingly stringent, i.e., one

© NCERT not to be republished

moves from a broad set of parameters to a narrow set, with the hope of identifying a small group of molecules that exhibits the desired property.

Virtual screening may include the use of: (a) general filters to identify drug like compounds with a desired Absorption, Distribution, Metabolism and Excretion (ADME) property, (b) Ligand based methods that encompass machine learning techniques, pharmacophore based search, and (c) structure based methods that include protein-ligand docking. Once a compound passes through these filters, one can use them for biological screening, synthesis, testing and so on.

Box 2

Common terminologies in cheminformatics

(1) **High Throughput Screening (HTS)**— A large scale automated process where millions of compounds are tested for a desired property.

(2) **Hits**—Activity observed during high-throughput screening, generally defined by percent activity of new compounds in comparison to well defined and known compounds.

(3) **False positive**—During screening, one may observe situations where a compound is found active in an assay but may turn out to be inactive towards a certain biological target.

(4) **Lead compound**— A compound that is biologically and pharmacologically active with desired properties, and that can be processed further.

(5) **Library**— An inventory of compounds that fulfill the criteria for screening against specific cellular targets.

(6) **New Chemical Entity**— A novel molecule discovered in the lab that has not yet entered clinical trials.

(7) **Off target activity**— Molecular interactions between chemical compounds and cellular molecules that do not bind the target.

SUMMARY

- Protein informatics is a growing field of information technology in which information about any protein is collected using sophisticated techniques. Raw data of proteins collected through various means is used to retrieve crucial information about the protein of interest.
- Primary structure of a protein can be analysed using ProtParam tool of ExPASy Proteomics Server. Isoelectric point, aliphatic Index, instability index and Grand Average

© NCERT not to be republished

Hydropathy (GRAVY) value of a protein are calculated using this server. Secondary structure of a protein is predicted using APSSP, CPSSP, SOPMA, and GOR.

- Homology modelling, fold prediction, and *de novo* protein structure prediction are common computational methods used to predict protein 3D structure.
- Cheminformatics combines the computational and informational techniques to understand the problems related to chemistry. The information used in cheminformatics includes information on physical properties, 3-D molecular crystal structures, chemical reaction pathways, etc.
- Pharmacophore modelling is a method which gives description about the molecular features that define molecular recognition of a ligand. Lipinski's rule of five (RO5) outlines key molecular properties of compounds which is helpful in choosing potential drug compounds.

EXERCISES

1. What is the role of information technology in determination of protein properties?
2. What type of protein raw data is used for computationally extracting information about the protein?
3. Name any two common tools for domain prediction.
4. What is the significance of cheminformatics?
5. Which of the following is not a rule in Lipinski's rule of five (RO5)?
   (a) No more than 10 hydrogen bond receptors
   (b) Partition coefficient logP of less than 5
   (c) Not more than 5 hydrogen bond donors
   (d) Molecular weight above 500g/mol
6. Which of the following properties of protein is not included in primary structure prediction?
   (a) Aliphatic index
   (b) Fold prediction
   (c) Instability index
   (d) Isoelectric point

© NCERT not to be republished

11150CH11

Chapter 11

Programming and Systems Biology

11.1 Programming in Biology

11.2 Systems Biology

11.1 Programming in Biology

From an era of manual computation, we are currently in a phase of large scale (i.e., high-throughput) data generation, automated analysis and prediction. Technological advancements have proven to be a boon for generating huge data, unthinkable a few decades ago. However, the arrival of massive data has also thrown massive challenges in the storage, visualisation, transfer, analysis and interpretation of data.

The task that looked gigantic a decade back appears trivial now.

The emergence of artificial intelligence (AI) and machine learning (ML) techniques has changed research practices in almost every field. It is increasingly evident that, in future, young biotechnology students working at the cutting edge of science may require basic programming knowledge and comfort with chemistry and statistical methods.

The purpose of this chapter is not to give an exhaustive description of programming languages but to offer a gentle introduction to some of the most popular high level languages relevant to biologists.

© NCERT not to be republished

Although bioinformatics software is being developed for all the available operating system (OS) platforms, majority of successful application have been developed on Linux platform. From the beginning of bioinformatics, PERL is always at the core of sequence based large data handling. Now a days these platforms are being enriched with the advanced performing language, normally Python and R provides strong facilities of statistical packages for solving biological problems. Similarly, Python modules are continuously being enriched with visualisation and analysis modules for handling of large data set on standalone, web server as well as cloud computing. Beyond these, MATLAB also includes very good platform for bioinformatics data analysis. Description of few of the most advanced languages, active in the area of bioinformatics, is given below:

Python: It is a high level programming general purpose language created by Guido van Rossum (1991). It is an object oriented programming interactive language that can run on unix, mac and windows. Python is very popular within bioinformatics community largely because: (i) of the clear meaning of terms used and the structure of statements (ii) it's expressivity and alignment to object-oriented programming, and (iii) the availability of libraries and third-party toolkits. Python has been successfully used for sequence and structure analyses, phylogenetics and so on.

R: The name R has been derived from its inventors, Robert Gentleman and Robert Ihaka, who developed this language. R language has gained wide acceptance as a rapid and reliable functional programming language that is ideal for high volume analysis, visualisation and simulation of biological data. The software is free and open source. The R language has been used for analysis of genome sequence and biomolecular pathways.

Moving from data analysis to designing the systems, new programming languages have emerged. Among them are — GEC (Genetic Engineering of living Cells), a rule based language developed by Microsoft and Kera, an object oriented knowledge based programming language developed by Dr. Umesh P. of University of Kerala. Kera (short form of Kerala, also means coconut) captures information on the genome, proteins and the cell, using a user edited biological library called *Samhita*.

© NCERT not to be republished

11.2 SYSTEMS BIOLOGY

11.2.1 Introduction

In order to understand the mysteries of nature, scientists are performing experiments since long. Findings of these experiments are recorded in the form of data in the literature. Starting from small bit of data to large ones, data are being collected from decades of experimental efforts. Presently, a large size of biological data is being generated and stored in the digital format in a variety of storehouses called databases. These digital data are the resources which make the foundation for researchers to develop such computational models which can perform tasks similar to our complex biological systems, i.e., those we observe in real *in-vitro*/*in-vivo* experiments or real life. Implementation of such ideas is being performed with mathematical and computational models to mimic complex biological systems. These models are called system models. Therefore, you can visualise the systems biology as the representation of system models. Now-a-days systems biology has become an area for intensive research with potent application. Thus, it is an interdisciplinary field of study that focuses on complex biological interactions within biological systems (Fig. 11.1). The concept of systems biology is being adopted in a variety of biological contexts, particularly from last two decades onwards. The human genome project is one of the most glorious seedings of a thought of systems biology, which led to new avenues of today's form of systems biology. Presently, systems biology models can provide theoretical description for discovery of emergent functional properties of cells, tissues and organisms, similar to those which were only possible through experiments. Examples of most efficient system models are metabolic or signaling network. Along with fundamental understanding of the mechanism of action of biological systems, systems biology is intensely being utilized into potent applicability, e.g., in the areas of health and diseases from biological networks to modern therapeutics.

11.2.2 Historical perspective

Before the emergence of systems biology, the scenario of research in biological sciences (e.g.,1900–1970) was wondering around physiology, population dynamics,

© NCERT not to be republished

enzyme kinetics, control theory, cybernetics, etc. as the segmental components of research. The systems biology has been mapped to be evolved from a physiological description, when in 1952 Alan Lloyd Hodgkin and Andrew Fielding Huxley (Nobel laureates) described a mathematical model for action potential propagation along the axon of a neuronal cell. More evolved implementation of theory emerged in 1960 when the first computer model of the heart pacemaker was developed by Denis Noble [PMID 13729365]. The systems biology was formally launched by systems theorist Mihajlo Mesarovic in 1966 at the Case Institute of Technology in Cleveland, Ohio, titled "Systems Theory and Biology". In 1968, first theory about systems biology was published by Ludwig von Bertalanffy, which is considered as precursor of this discipline. The duration betwen 1960s and 1970s was the decade of development of multiple aspects of complex molecular systems, such as the metabolic control analysis and the biochemical systems theory. Furthermore, skepticism of systems theory with molecular biology was broken by the development of theoretical biology, which includes the quantitative modelling of biological processes. Since 1990s, functional genomics is generating large quantities of high-quality biological data, which are helping in the development of more realistic models. In continuation of these developments in the area of systems biology, National Science Foundation (NSF) put forward a challenge to mathematically model the whole cell. In this direction, in 2003, Massachusetts Institute of Technology (MIT) started search of solution of this challenge in association with CytoSolve. Finally, in 2012 whole cell model of *Mycoplasma genitalium (cell wall less bacterium)*, for prediction of cell viability in response to the genetic mutations, was developed by Mount Sinai School of Medicine, New York. Presently, a big systems biology project, namely 'Physiome' is available (http://physiomeproject.org/). This project is aimed at developing a multi-scale modelling framework for understanding the physiological function that allows models to be combined and linked in a hierarchical fashion. For example, electromechanical models of the heart, need to be combined with models of ion channels, myofilament mechanics and signal transduction pathways at the subcellular level and then to link these processes

© NCERT not to be republished

to models of tissue mechanics, wavefront propagation and coronary blood flow—each of which may well have been developed by a different group of researchers.

11.2.3 Theme behind the systems biology

To cover diverse disciplines of biology, systems biology has been observed from different aspects. The reductionist worked on identification of components and interactions of a system, but no convincing method could be evolved to describe the pluralism of system. Pluralism can be better observed through quantitative measures of multiple components simultaneously and this can only be possible by mathematical models containing rigorous data integration. In this way, it can be said that systems biology is the observation of system by integrating different components together (Fig.11.1). Covering all the individual components together at the core of theme of systems biology is: 'Object network mapping and its integration with interdependent dynamic event—kinetics with partial differential equations'.

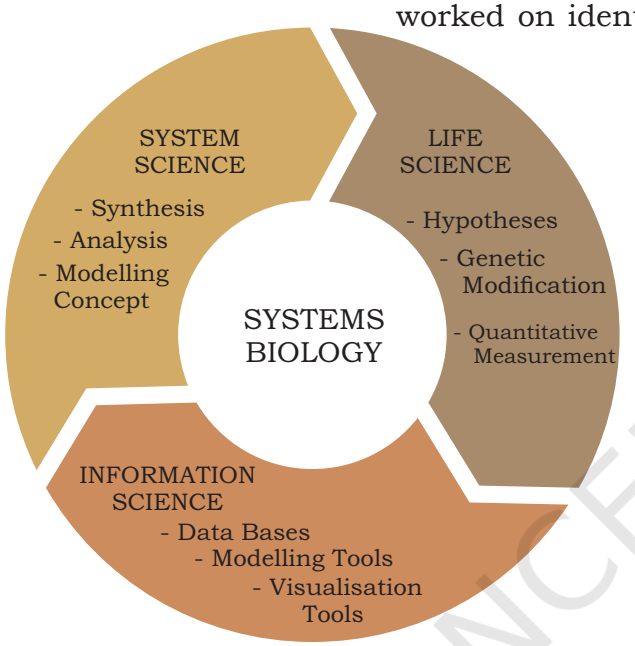


Fig. 11.1: Depiction of Systems Biology as an interdisciplinary field of study that focuses on complex biological interactions within biological systems

11.2.4 Protocol for systems biology experiments

To perform a standard systems biology experiment, discrete steps, as shown in Fig.11.2, are followed.

The whole protocol basically involves definition of problem, designing of experiment, execution of the experiments to generate data, collection of the resultant data and their arrangement in appropriate file formats followed by the development of network interface. This is followed by transfer of this network interface which should be precise as well as mechanism based so that the model can be developed accordingly. This is further followed by analysis of discrepancies between model based simulation results and experimental data and accordingly model the hypothesis with reference to the discrepancies observed. Finally, the simulation is repeated and tested again and

© NCERT not to be republished

again, and new hypotheses are incorporated into the model.

Thus, the work flow of computation for systems biology (as depicted in Fig. 11.2) requires data-management, optimisation of network development parameters, performance analysis and evaluation.

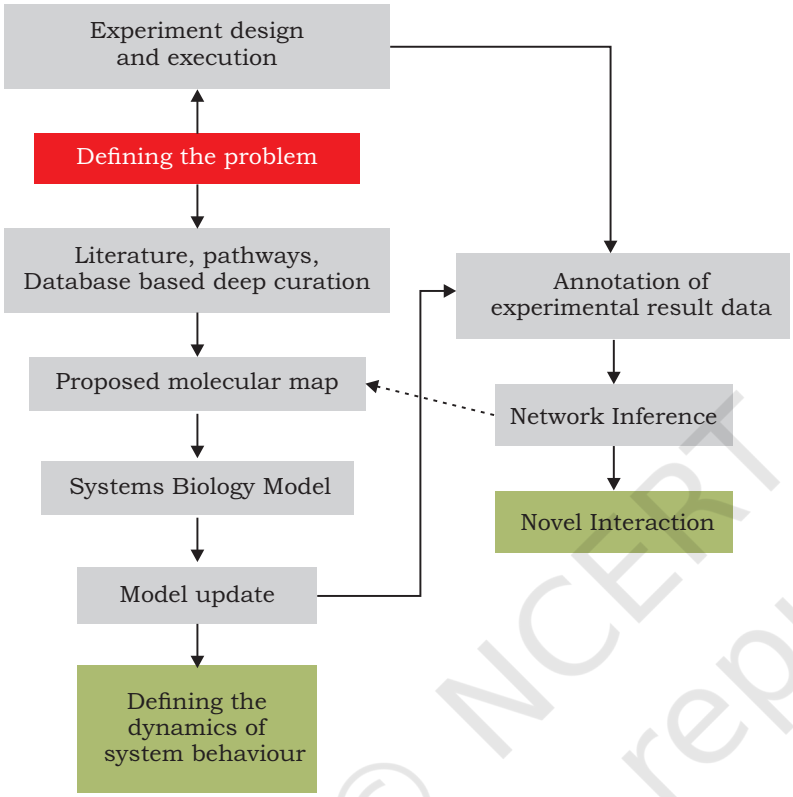


Fig. 11.2: Workflow for implementation of systems biology experiment

Standards for data management have been defined for collection of structure data for systems biology. Accordingly, three basic aspects are considered for data management which are explained below—

(i) Minimum information

Minimum information represents a set of essential supporting information needed from different experiments as microarray, proteomic, biological, and biomedical investigations. Incorporation of metadata about these collected data is an important point to care.

(ii) File formats

The collected data for minimum information are stored in specific file formats. These formats are generally

© NCERT not to be republished

Extensible Markup Language (XML) based, which has facility to be automatically processed by computers.

(iii) Ontologies

Ontologies define a semantic annotation of data, which represents the hierarchical relationship between different terms. Few important examples are, **Gene Ontology (GO)** and **Systems Biology Ontology (SBO)**.

Current data-management systems include spreadsheets, web-based electronic lab notebooks (ELN), and laboratory information management systems (LIMS). The data-management systems have been customised in such a way that they can be accessed and integrated with different analysis tools and computational workflows. Systems like Konstanz Information Miner (KNIME), caGrid23, Taverna24, Bio-STEER25 and Galaxy26, allow the construction, execution and sharing of specialised workflows. These workflows provide computational pipeline by enabling data exchange, data integration and inter-tool communication. A list of data management, network inference, curation, simulation, model analysis, moleular interaction, and physiological modelling tools have been given in Table 11.1.

Table 11.1 A resource matrix of software, tools and data resources

Facilities	Tools / Software
Data management	Taverna, MAGE-TAB, Bio-STEER, caGrid
Network inference	MATLAB, R, BANJO
Curation	CellDesigner, PathVisio, Jdesiner
Simulation	MATLAB, CellDesigner, insilico IDE, ANSYS, JSim
Model analysis	MATLAB, BUNKI, COBRA, NetBuilder, SimBoolNet
Moleular interaction	AutoDock Vina, GOLD, eHiTS
Physiological modelling	PhysioDesigner, CellDesigner, OpenCell, FLAME

These systems modelling tools include set of inter-connected partial differential equations (PDEs) which represent spatiotemporal systems. The PDEs are solved by the Finite Element Method (FEM), which is a numerical technique for approximate solutions for PDEs. PDEs can be solved by ANSYS, FreeFEM++, OpenFEM and MATLAB.

© NCERT not to be republished

There are several tools, which are used for systems modelling. These include: JSim, OpenCell and Flexible Large-scale Agent-based Modelling Environment (FLAME) etc. Many other simulation tools are under development which touch more real life aspects of simulations.

11.2.5 Model-analysis methods

Several mathematical techniques have been developed to analyse the behaviour of complex biological models. Some basic principles for model analysis are presented below—

(i) Sensitivity analysis

Sensitivity analysis describes about the stability and controllability of system against various distractions. Some of the important tools for sensitivity analysis are: SBML-SAT, MATLAB SimBiology, ByoDyn and SensSB.

(ii) Bifurcation and phase-space analysis

Bifurcation and phase-space analysis is performed to analyse the system model to discover the possible steady as well as dynamical tendencies. Some of the important tools are: AUTO, XPPAut, BUNKI and ManLab.

(iii) Metabolic control analysis

Metabolic control analysis (MCA) is performed for understanding the relationship between the properties of a metabolic network (at steady state) and component reactions. The MetNetMaker is a tool for this.

© NCERT not to be republished

SUMMARY

- With increasing amount of data produced everyday by biologists, it has become all the more important to competently handle complex datasets in order to generate and explore hypotheses. Programming languages make it easier for the scientists to access, filter and manipulate the biological data.
- Some of the most advanced programming languages include Python and R. Python can run on unix, mac and windows, and is used for visualisation and analysis of sequence and structure datasets. The R language provides facilities of statistical tools, and is suitable for high volume analysis and visualisation.

- Systems biology utilises computational methods to analyse complex biological datasets. Examples of systems models are metabolic and signaling network.
- Data management, optimisation of network development parameters, performance analysis and evaluation are some of the important requirements of computation for systems biology. There are three aspects of data management in systems biology, namely minimum information, file formats and ontologies.

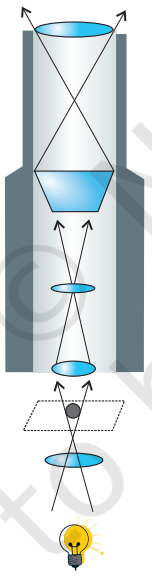
EXERCISES

1. Why are programming languages a boon for biologists?
2. Name the various aspects of data management for systems biology.
3. Choose the INCORRECT statement
   (a) Python is a programming language.
   (b) Biologists do not require knowledge of statistical tools for handling datasets.
   (c) Most of the applications have been developed on Linux platform.
   (d) Python provides third-party toolkits.
4. Systems biology is
   (a) the systematic study of all the living organisms.
   (b) the thorough study of all biochemical and signaling pathways.
   (c) the detailed study of biological systems through computational and experimental methods.
   (d) the study of dynamics of enzymes.
5. Which of the following is NOT included in data management systems?
   (a) Metabolic control analysis
   (b) Spreadsheets
   (c) Web-based electronic lab notebooks (ELN)
   (d) Laboratory information management systems (LIMS)
6. What is the need of systems biology?
7. What is the fundamental difference between systems biology and physiology?
8. Explain systems biology as collection of approaches and tools.
9. Is systems biology cell-centric?

© NCERT not to be republished

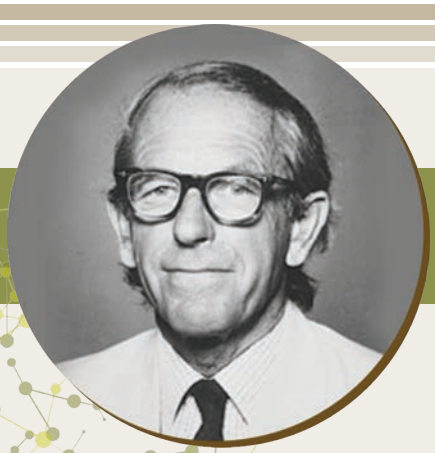
Unit V
Tools and Techniques: Basic Concepts

Chapter 12
Tools and Techniques



Considering the fact that biotechnology is an experimental science and involves a lot of experimentations; therefore, research in this field depends highly on sophisticated laboratory methods. Advances in biotechnology were closely followed by the development of newer tools and techniques in biological sciences. These new methods opened new avenues for research and investigation in the field of biotechnology. It is, thus, important to appreciate the experimental tools available to biotechnologists in order to understand the progress and future directions of this rapidly moving area of science. Some of the important experimental methods including methods of cell and molecular biology will be discussed in this unit.

© NCERT not to be republished



Frederick Sanger (1918-2013)

Frederick Sanger (1918–2013) was a British biochemist and molecular biologist who had two Nobel Prizes in Chemistry to his credit. He was awarded the first Nobel Prize in 1958 for the discovery of structure of insulin molecule, and second Nobel Prize in 1980 for his work (in collaboration with Paul Berg and Walter Gilbert) on the determination of base sequences of nucleic acids. He is, by far, the most influential biochemist in history. His technique of deciphering DNA sequences was based on 'read-off' methods using acrylamide gel. In 1977, Sanger sequenced the genome of bacteriophage ΦX174, the first genome to be completely sequenced. Most of his later contributions laid the foundation of molecular biology and are being utilised in every biotechnology application.

© NCERT not to be republished

11150CH12

Chapter 12

Tools and Techniques

12.1 *Microscopy*
12.2 *Centrifugation*
12.3 *Electrophoresis*
12.4 *Enzyme-linked Immunosorbent Assay (ELISA)*
12.5 *Chromatography*
12.6 *Spectroscopy*
12.7 *Mass Spectrometry*
12.8 *Fluorescence in situ hybridisation (FISH)*
12.9 *DNA Sequencing*
12.10 *DNA Microarray*
12.11 *Flow Cytometry*

12.1 Microscopy

Biological studies and explorations cannot be imagined without a microscope as it enables us to see something which is beyond the scope of our eyes. Today, the technique of microscopy has become so much advanced that a researcher can not only see a highly magnified image of a very minute structure but also can visualise the three dimensional structure of such objects. Using powerful electron microscopic techniques, even the DNA molecule of bacteria and viruses have been visualised.

The use of first microscope dates back to 1665 when the British physicist Robert Hooke designed a simple microscope using combination of magnifying lenses (Fig. 12.1) and observed the slices of cork, and coined the term '*Cellulae*' or '*cell*' to that honeycomb like structure. You are aware that Matthias Jacob Schleiden and Theodor Schwann proposed cell theroy in 1838 on the basis of observation of cells in plants and animals.

© NCERT not to be republished

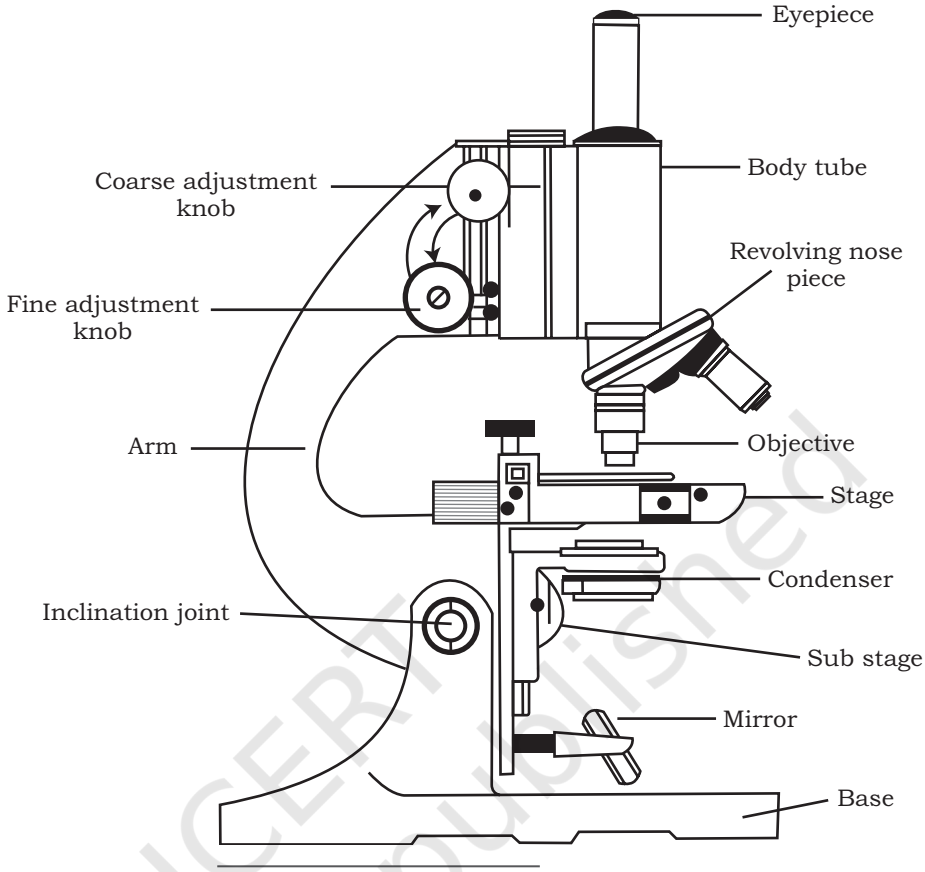


Fig. 12.1: Microscope

12.1.1 Magnification and Resolution

Let us now focus on the principle on which the technique of microscopy is based. Two optical properties are extremely important for an optical instrument like microscope. One is the power to magnify and the other is the ability to resolve.

Magnification or magnifying power of a microscope is the ability by which the retinal image size can be increased. Thus in simple terms magnification is—

$$\frac{\text{Size of retinal image with the help of microscope}}{\text{Size of the retinal image without using microscope}}$$

You may have studied in physics that magnification (M) of a lens is measured as per the following formula (in which f is focal length of the lens and d is the distance of object from the lens).

$$M = \frac{f}{f - d}$$

© NCERT not to be republished

Normally, the microscope which is used in the laboratory is a compound microscope in which two sets of lenses are there. One is called objective lens, which remains close to the object to be seen, and the other is eyepiece through which the observer sees. It is needless to mention that the object, objective lens, eyepiece, and the eye of the observer have to be in the same line for the passage of light to see the magnified image of the object. In simple words, magnification of a microscope is product of the magnifying power of the objective lens and the magnifying power of the eyepiece ($M_o \times M_e$).

Resolving power is another important property of a microscope, which is the ability to form separate images of the two objects situated very close to each other. It can be measured by the smallest distance between two points.

12.1.2 Functioning of a light microscope

You have already studied about the structure of a compound microscope in previous class, yet to recapitulate, as you can see in the Fig. 12.1, a compound microscope consists of a base on which a stage is fitted with a central hole. Attached to the base is an arm to which a body tube is fitted in such a way that it aligns with the hole of the stage. At the lower end of the body tube, a nose piece is fitted on which two to four objective lenses may be present. By rotating the nose piece, one of the objective lenses can be placed above the hole present on the stage where object to be seen is placed on a glass slide. At the upper end of the body tube, an eye piece is fitted through which an observer can see under the microscope. There are adjustment screws (coarse and fine) on the arm which facilitate in adjusting the distance of objective lens from the object present on the stage. Below the stage, there is a source of light (which may be a reflective mirror or a bulb to illuminate the object and facilitate the formation of image through objective lens and eyepiece) (Fig. 12.2). In addition, there is a condenser present between the light source and the stage, which is important for focusing light on the object. You might have observed that both objective lenses and eye pieces are of different magnifying power. In a student microscope the eyepiece has the magnifying power of 10× or 15× and that of the different objective lenses fitted on the nose piece are of 4×, 10×, 40/45× and 100×. The technique of microscopy

© NCERT not to be republished

just discussed is also called **bright field microscopy** as light is used to illuminate the object to be seen. Therefore, in order to distinguish different regions of the object, the same is stained with specific dyes or stain. Carmine, eosin, safranin, methylene blue, giemsa, etc., are few such stains commonly used for light microscopy.

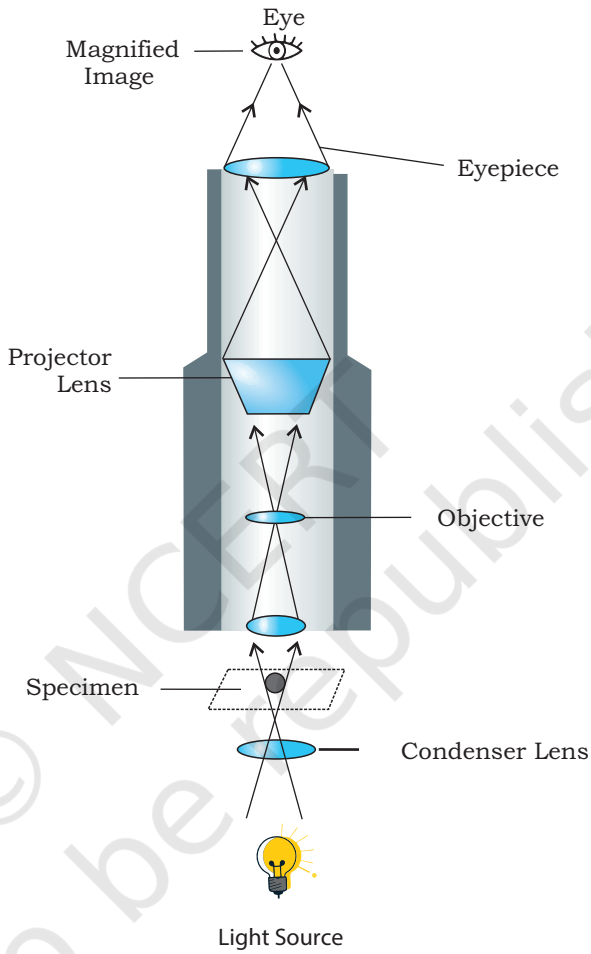


Fig. 12.2: Pathway of light in a light microscope

12.1.3 Different forms of microscopy

Studying minute details of internal organisation of tissues/ cells is so diverse that it cannot be achieved by light microscopy only. Therefore, by making one or the other kind of maneuvering, quite diverse forms of microscopy are used. In one such maneuvering, light falling on the object from the central condenser is blocked by a disc and illumination of the object is done by an oblique light beam, which is reflected off from the slide and the image

© NCERT not to be republished

is illuminated against the dark background. Therefore, such a microscopy is known as **Dark Field Microscopy**. Mitochondria, nuclei, vacuoles, etc., can be easily detected using this. Similarly, in a different form called **Phase Contrast Microscopy,** wave amplitude and phase of light passing through the transparent object is changed. This change depends on the density of the part of the object or specimen. Such a change is more in the area where density is comparatively high and as a result of which, varied contrast of different regions of the object can be seen. This is especially helpful in the study of cell organelles and chromosomes. Staining of the object or specimen with some specific dye is routinely done. There are some special types of dyes e.g., acridine orange, bisbenzimide, merocyanine (also called fluorophores). These dyes are capable of emitting light of longer wavelength after being illuminated, a property called **fluorescence**. As a result of this, the fluorophore stained object looks more illuminated and of different colour depending on the dye used. In **fluorescence microscopy**, the same principle is applied. Object to be seen is stained with fluorophore to study a specific part of organelle or molecule. After illuminating the object under fluorescence microscope, the specifically stained region is easily seen or observed. This is helpful in identifying bacteria or viruses to know the cause of infection and immunodiagnosis.

Electron microscopy is a highly sophisticated technique in which the object to be studied is bombarded with electron beam which is approximately 1,00,000 times shorter in wavelength than visible light. The electron beam in an electron microscope magnifies the image with the help of electromagnetic lenses. The entire passage of electron is in vacuum and the generated image is viewed on a fluorescent screen and not through eyepiece. Owing to the very shorter wavelength of the electron, the image produced by an electron microscope is of very high resolution. Two types of electron microscopy are used; transmission electron microscopy and scanning electron microscopy. In **Transmission electron microscopy,** the ultra-thin heavy metal salt (of lead, tungsten, etc.) coated section of the object or specimen is placed in such a way that the electron beam passes through it to create the image. In the other technique of electron microscopy,

© NCERT not to be republished

reflected electron beam from the gold or platinum coated surface of the object creates the image. In this technique, highly magnified and resolved image of the object surface is generated, therefore, it is called **Scanning electron microscopy**.

In last few decades, yet another more sophisticated microscopic imaging technique has been developed and used called the **confocal microscopy**. Confocal microscopy is useful in resolving detailed structures within fixed cells/tissues and gives sharp images of the objects. To examine an object using confocal microscopy, it is first fluorescently labelled and then analysed under a confocal microscope in high resolution.

12.2 CENTRIFUGATION

You have studied about various biomolecules like proteins, nucleic acids, etc., present in cells of all living organisms. To study these biomolecules, you need to isolate them by using one or the other separation techniques. Centrifugation is one such technique in which particles or molecules are separated based on their densities under the influence of gravitational force (g), by spinning them in a solution around an axis at high speed using **centrifugal force**. The equipment used is called centrifuge (Fig. 12.3), which is of different types depending on its use. It consists of a base, a rotating container (spinning vessel/rotor) and a lid. The

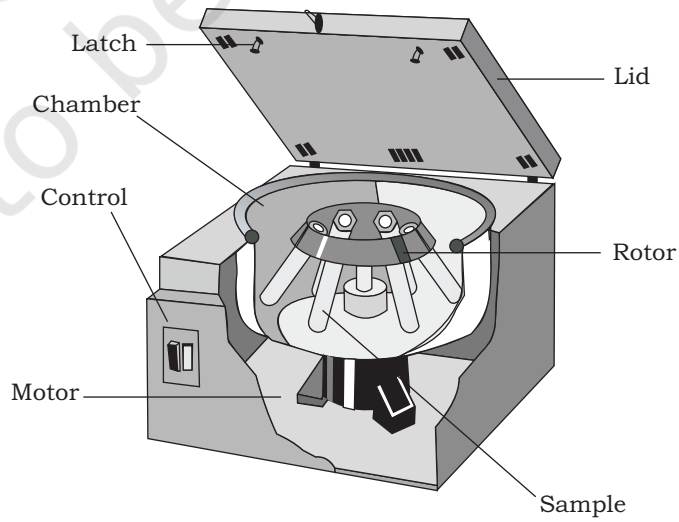


Fig. 12.3: Basic structure of centrifuge

© NCERT not to be republished

spinning vessel contains several centrifuge tubes. The cell extract or mixture is taken in centrifuge tubes and allowed to spin at a desired speed (revolution per minute; rpm) for a specific period of time leading to settlement of particulate material at the bottom of the centrifuge tubes.

12.2.1 Types of Centrifuge

Based on the principle and application, following types of centrifugations are performed—

Differential Centrifugation—It is based upon the differences in the sedimentation rate (centrifugal force) of particles of different size and density. It is used to separate large cellular structures, the nuclear fraction, mitochondria, chloroplasts, or large protein.

Density-gradient Centrifugation— In order to separate biological particles of similar size but differing in densities, one can use density gradient centrifugation. In this type of centrifugation, a density gradient is developed in centrifuge tubes. Depending upon their densities, different molecules get sedimented at different levels. The heavier molecules move outward and lighter ones remain in inner part in the centrifuge tubes. The greater the difference in density, the faster they move.

Ultracentrifugation— When centrifugation is carried out at very high speed, i.e., 100,000 x/g or more to separate the molecules, it is called **ultracentrifugation**. Commonly used centrifuges are:

- Table top/clinical centrifuge or microfuge
- High-speed centrifuge
- Ultracentrifuge

12.3 ELECTROPHORESIS

Electrophoresis is a method of separation on the basis of charge to mass ratio of macromolecules under the influence of an electric field. Electrophoresis is a Greek word meaning 'to bear electrons'. The prefix *electro* refers to electricity which is required to migrate molecules and the suffix *phoresis* means 'migration' or 'movement'. It was observed for the first time in 1807 by Russian professors

© NCERT not to be republished

Peter Ivanovich Strakhov and Ferdinand Frederic Reuss. They noticed the migration of the clay particles dispersed in water in the presence of constant electric field.

Principle

Many important biological molecules such as nucleotides, DNA, RNA, peptides and proteins bear ionisable groups, and therefore at any given pH exist in solution as electrically charged species either as cations or anions. Under the influence of an electric field, these particles will migrate either towards cathode or anode, depending on their net charge.

The mobility of a molecule is inversely proportional to its size and directly proportional to its charge, allowing them to be separated from one another.

12.3.1 Agarose Gel Electrophoresis

In this type of electrophoresis, the gel is a matrix of agarose molecules that are held together by hydrogen bonds and form tiny pores. Gels for DNA separation are often made out of a polysaccharide called agarose, which comes as dry and powdered flakes. When the agarose is heated in a buffer and allowed to cool, it will form a solid and slightly squishy gel.

Gel is a slab of jelly-like material, placed in a gel box. One end of the box is connected to a positive electrode, and another end with a negative electrode. The gel box is filled with a salt-containing buffer solution that can conduct current. The end of the gel with the wells is positioned towards the negative electrode. The other end of the gel is positioned towards the positive electrode to which the DNA fragments will migrate (Fig. 12.4).

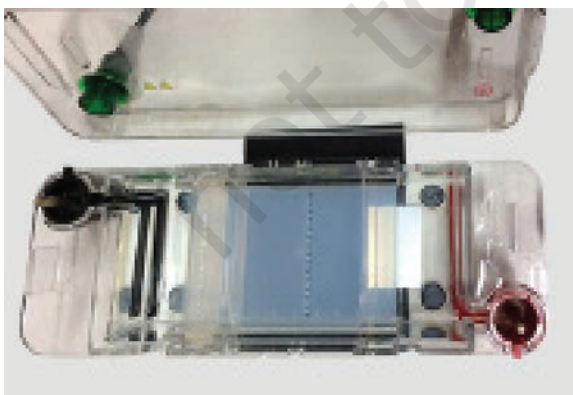


Fig 12.4: Agarose gel electrophoresis unit to separate nucleic acid

DNA molecules are negatively charged. Gel electrophoresis of DNA fragments separates them on the basis of size only. Using electrophoresis, we can check the different DNA fragments present in a sample and determine their absolute size with the help of **DNA ladder** made up of DNA fragments of known sizes.

When the power is turned on, the current begins to flow through the gel.

© NCERT not to be republished

DNA molecules have a negative charge due to the presence of phosphate groups in their sugar-phosphate backbone; therefore, they move through the matrix of the gel towards the positive electrode (anode).

The voltage for running an agarose DNA gel lies in the range of 80–120 V. As the electric current is applied, shorter pieces of DNA travel through the pores of the gel matrix faster than longer ones. Thus the longest pieces of DNA remain near the wells while the shortest pieces of DNA are close to the positive end of the gel.

12.3.2 Visualising the DNA fragments

The equipment used for visualisation of target DNA is ultra-violet (UV) trans-illuminator. Ethidium bromide (EtBr) is most commonly used for staining DNA. This stain can be mixed in the gel mixture, in the electrophoresis buffer or the gel is stained after it is run. Molecules of the EtBr intercalates within DNA bases and fluoresce under UV light. Despite its advantages, ethidium bromide is a potential carcinogen, so it must be handled with great care (Fig. 12.5).

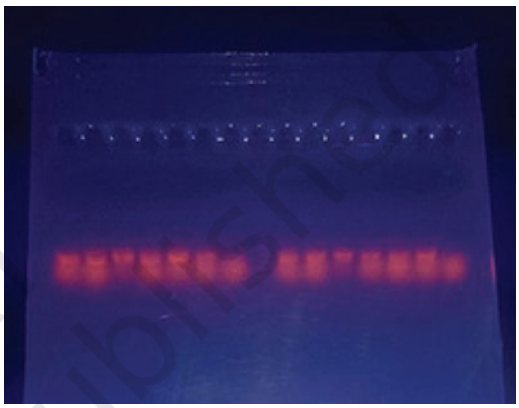


Fig. 12.5: Visulalisation of DNA bands under UV light

12.3.3 Polyacrylamide Gel Electrophoresis (PAGE)

PAGE is an analytical method used to separate components of a protein mixture based on their size. To provide uniform charge to the protein molecules, an anionic detergent called sodium dodecyl sulfate (SDS) is used to bind the proteins and give them negative charge. Proteins are then separated electrophoretically according to their size using a gel matrix made of polyacrylamide in an electric field.

Polyacrylamide is produced as a result of the polymerisation reaction between acrylamide and N,N'-methylene-bis-acrylamide (BIS) using a catalyst. The degree of polymerisation or cross-linking can be controlled by adjusting the concentration of acrylamide and BIS. The more the cross-linking, the harder the gel. Hardness of the gel, in turn, modulates the friction experienced by macromolecules when they travel through the gel during PAGE, thus affecting the resolution of separation. Loose gels (4–8% acrylamide) allow higher molecular weight molecules

© NCERT not to be republished

to migrate faster through the gel while hard gels (12–20% acrylamide) restrict the migration of large molecules and selectively allow small ones to move through the gel.

12.3.4 Tracking dye

As DNA, RNA, and proteins are mostly colourless, their progress through the gel during electrophoresis cannot be easily followed. Anionic dyes of a known electrophoretic mobility are therefore, usually included in the sample buffer. A very common tracking dye is bromophenol blue. This dye is coloured at alkali and neutral pH, and is a small negatively charged molecule that moves towards the anode and it is coloured at alkali and neutral pH. Being a highly mobile molecule it moves ahead of most proteins and nucleic acids. As it reaches the anodic end of the electrophoresis medium, electrophoresis is stopped. Other common tracking dyes are xylene cyanol, which has lower mobility, and Orange G, which has a higher mobility.

Visualisation of protein on gel— Coomassie Brilliant Blue R-250 is the most popular protein stain. It is an anionic dye, which non-specifically binds to proteins.

Box 1

Surprise Discovery of Jelly property

A Japanese Emperor and his Royal party were lost in the mountains during a snow storm and arrived at a small inn; they were treated by the innkeeper with a seaweed jelly dish with their dinner. Maybe the innkeeper prepared too much jelly or the taste was not so palatable but some jelly was thrown away, freezing during the night and crumbling afterwards by thawing and draining, leaving a cracked substance of low density. The innkeeper took the residue and, to his surprise, found that by boiling it up with more water, the jelly could be remade.

Agar

One of the scientists named Koch used to culture bacteria on the sterile surfaces of boiled cut potatoes. This was unsatisfactory because bacteria would not always grow well on potatoes. He then tried to solidify regular liquid media by adding gelatin but it was digested by many bacteria and melted when the temperature rose above 28°C. A better alternative was provided by Fanny Eilshemius Hesse, the wife of Walther Hesse, one of Koch's assistants. She suggested the use of agar as a solidifying agent—she had been using it successfully to make jellies for some time. Agar was not attacked by most bacteria and did not melt until reaching a temperature of more than 42°C. Agarose is one of the two principal components of agar, and is purified from agar by removing agar's other component, agaropectin.

© NCERT not to be republished

Proteins in the gel are fixed by methanol and acetic acid and simultaneously stained. The excess dye incorporated into the gel can be removed by destaining with the same solution without the dye. The proteins are detected as blue bands on a clear background. After the visualisation by a staining (protein-specific) technique, the size of a protein can be calculated by comparing its migration distance with that of a known molecular weight ladder.

12.4 Enzyme-linked Immunosorbent Assay (ELISA)

Enzyme-linked immunosorbent assay (ELISA) was invented by two Swedish scientists, Eva Engvall and Peter Perlman in 1971. ELISA is a quantitative method used for the measurement of antigen and antibody concentration in a given sample. This is done by monitoring the antigen-antibody interaction with the help of an enzyme catalysed reaction. This detection system (an enzyme-conjugate) is covalently linked to a specific antibody that recognises a target antigen. The intensity of the color produced is detected by ELISA reader or spectrophotometer. ELISA is a safer and less costly assay as compared to many other immunological assays.

A number of modifications of ELISA have been developed, allowing qualitative detection or quantitative measurement of either antigen or antibody. These different types of ELISA can be employed qualitatively to detect the presence of antibody or antigen. Using the known concentrations of antibody or antigen, a standard curve is prepared to determine the unknown concentration of a sample.

Direct ELISA— In direct ELISA, the antigen or sample is coated on the microtiter plate wells and the enzyme-conjugated antibody 'directly' binds to the antigen. The enzyme linked to the antibody reacts with its substrate to produce a colourful product that can be measured using spectrophotometer/ELISA reader. The direct ELISA is faster and less error prone and does not require a secondary antibody. Disadvantage of direct ELISA is non-specific binding of antibody due to cross-reactivity.

Indirect ELISA— The indirect ELISA detects the presence of antibody in a sample in two stages (Fig. 12.6). Firstly,

© NCERT not to be republished

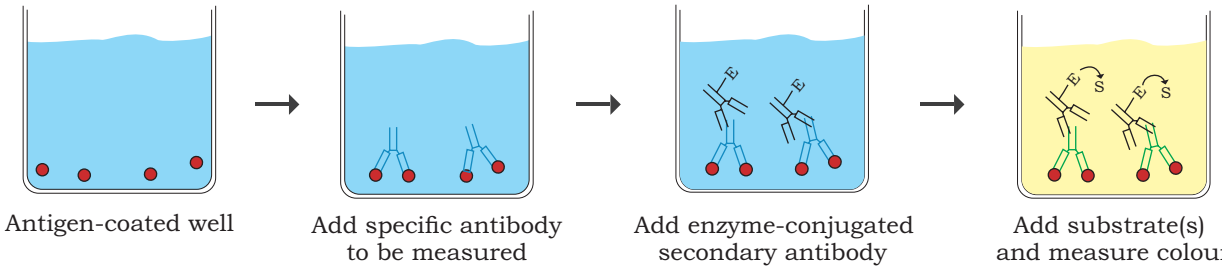


Fig. 12.6: Indirect ELISA

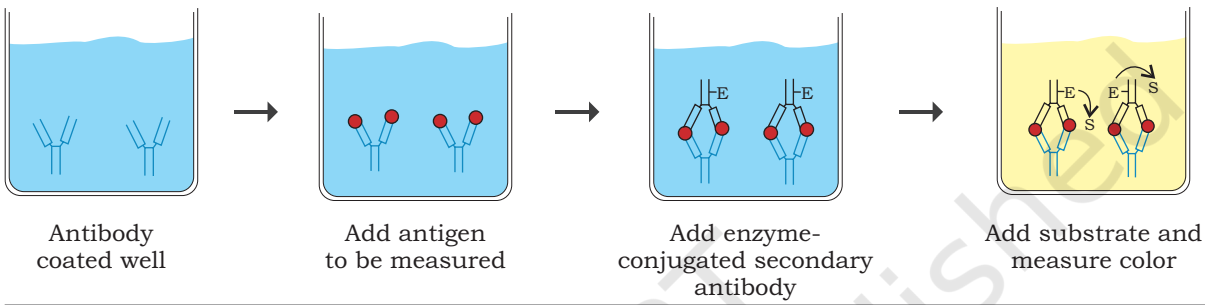


Fig. 12.7: Sandwich ELISA

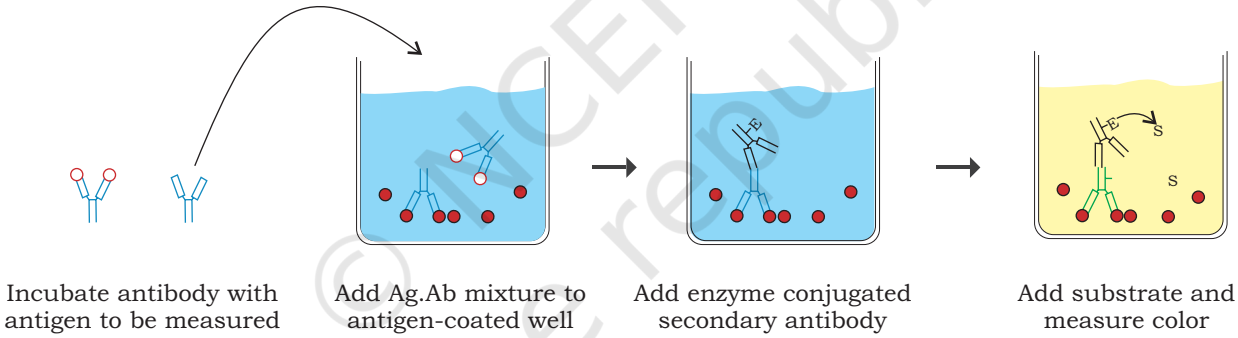


Fig. 12.8: Competitive ELISA

an unlabelled primary antibody (Ab1) is applied to antigen coated microtiter wells. The unbound extra primary antibody (Ab1) is then washed off. Next, an enzyme-conjugated secondary antibody (Ab2) specific for primary antibody (Ab1) is added. Any unbounded secondary antibody (Ab2) is washed off and the substrate is added. The amount of coloured product formed can be measured using spectrophotometer/ELISA reader. Indirect ELISA has increased sensitivity since more than one labelled antibody is bound per primary antibody.

Sandwich ELISA—In sandwich ELISA, the antibody is coated on the microtiter plate and is referred to as captured antibody. Using sandwich ELISA technique, the

© NCERT not to be republished

antigen can be detected or measured (Fig. 12.7). After the captured antibody is coated on the plate, antigen is added and allowed to react with the captured antibody. Excess antigen is washed off and then a second enzyme-conjugated antibody specific for a different epitope (the part of an antigen molecule to which an antibody attaches) on the antigen is loaded. The excess unbound enzyme-conjugated antibody is washed off and substrate is added and the coloured product produced is then measured using spectrophotometer/ELISA reader. High specificity and no requirement of antigen purification or limiting antigen are the main advantages of sandwich ELISA.

Competitive ELISA— The competitive ELISA helps in measuring amounts of antigen (Fig. 12.8). In this method, the antibody and antigen are first incubated in solution. In this step, the antibody is present in excess to the antigen; it will bind to its antibody in a concentration dependent manner leaving unbound antibody, accordingly. This antigen-antibody complex is then added to an antigen coated microtiter well. Thus, if more antigens are present in the sample, less free antibody will be available to bind to the antigen-coated well and *vice versa*. Finally, the enzyme-conjugated secondary antibody (Ab2), specific for the primary antibody is added to the plate and these bind to the primary antibody bound to the antigen on the plate. As in the case of indirect ELISA, addition of an enzyme-conjugated secondary antibody (Ab2) specific for the primary antibody helps in determining the amount of primary antibody bound to the well. Thus, in the competitive ELISA, the higher the concentration of antigen in the sample; the lower will be the free antibody and accordingly, the intensity of the developed colour will be less.

Application of ELISA

1. Presence of antigen or antibody in a sample can be estimated.
2. Concentration determination of serum antibody in a virus test can be carried out.
3. Detection of potential food allergens in food industry.
4. Disease outbreaks can be tracked down e.g., HIV, bird flu, cholera, etc.

© NCERT not to be republished

12.5 CHROMATOGRAPHY

Principle

Chromatography is a separation procedure for isolating various components of a mixture. The whole operation is time dependent, and involves a mobile phase and a stationary phase. The mobile phase may be a solution containing solutes to be separated and the eluent (liquid) that carries the solution through the stationary phase. Stationary phase may be adsorbent, ion-exchange resin, porous solid, or gel. The basis of chromatography is different migration properties of the solute molecules during passage through the stationary phase. Each solute in the original solution moves at the rate proportional to its relative affinity for the stationary phase. The stationary phase is usually packed in a cylindrical column. Fig. 12.9 reveals an outline description of the separation of three solutes from a mixture through chromatography.

The various components of a mixture, i.e., the solute molecules travel with the eluent molecules at different speeds

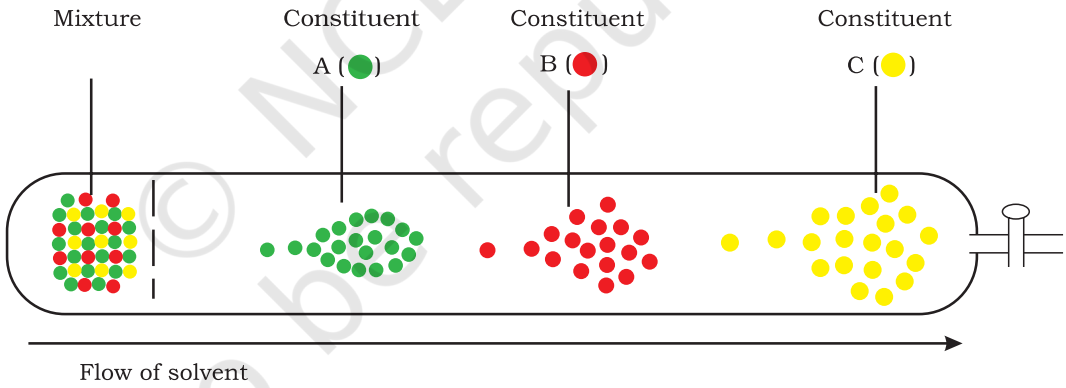


Fig. 12.9: Chromatography—a schematic description

depending upon their relative affinities for the resin particles. As a result, they separate and appear for collection at the other end of the column at different time intervals. The pattern of solute peaks emerging from a chromatography column is called a **chromatogram.** A typical HPLC chromatogram is shown in Fig. 12.10.

Liquid chromatography is used both as a laboratory method for sample analysis and as a preparative technique for large-scale purification of biomolecules.

© NCERT not to be republished

In recent year there have been rapid developments in the technology of liquid chromatography aimed at the isolation of recombinant products from genetically engineered organisms. Chromatography is a high resolution technique and therefore suitable for the recovery of high-purity therapeutics and pharmaceuticals. Chromatographic

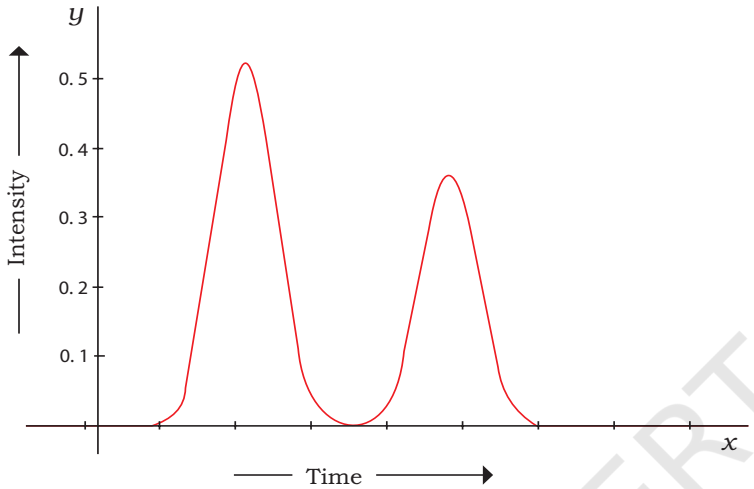


Fig. 12.10: HPLC chromatogram

methods are available for purification of proteins, peptides, amino acids, nucleic acids, alkaloids, vitamins, steroids and many other biological materials. These methods differ in the principles by which molecules are separated in the chromatography column.

Adsorption Chromatography (ADC): This chromatography is based on the adsorption of solute molecules onto polar adsorbents such as silica gel, alumina, diatomaceous earth and charcoal. Because the mobile phase is in competition with solute for adsorption sites, solvent properties are also important. Polarity scales for solvents are available to aid mobile-phase selection.

Liquid-liquid Partition Chromatography (LLC)—Partition chromatography relies on different partition coefficients (solubility) of solute molecules between two immiscible solvents. This is achieved by fixing one solvent as the stationary phase and passing the other solvent containing solute over it, a mobile phase. The solvents make intimate contact allowing multiple extractions of solute to occur.

© NCERT not to be republished

Ion-exchange chromatography (IEC)—The basis of separation in this procedure is the adsorption of solute ions on ion exchange resin particles on the column packing by electrostatic forces. Ion exchange chromatography may provide high resolution of macromolecules. Commercially, it is used for fractionation of antibiotics and proteins. Column packings include silica, glass and polystyrene; carboxymethyl and diethylaminoethyl groups attached to cellulose, agarose or dextran provide suitable resins for protein chromatography. Solutes are eluted by changing the pH or ionic strength of the liquid phase.

Gel Filtration Chromatography—This technique is also known as **molecular sieve chromatography, size exclusion chromatography** and **gel-permeation chromatography.** This is based on the size of molecules to be separated. Solute molecules present in the solution are separated in a column packed with gel particles of distinct porosity. Mostly, the cross-linked dextrants, agaroses and polyacrylamide gels are used for the packing of column. The penetration of the solute molecules through the column depends on the shape of solute molecules and their effective molecular size. Gel filtration can be used for separation of proteins and lipophilic compounds. Large-scale gel filtration columns are operated with upward-flow elution.

Affinity Chromatography (AFC): This separation technique exploits the binding specificity of biomolecules. Enzymes, hormones, receptors, antibodies, antigens, binding proteins, lectins, nucleic acids, vitamins, whole cells and other components capable of specific and reversible binding are amenable to highly selective affinity purification. Column packing is prepared by linking a binding molecule called a ligand to column, only solutes with appreciable affinity for the ligand are retained.

High-pressure /performance Liquid Chromatography (HPLC): It is based on general principles of chromatography. With the development of HPLC, the particle size of stationary phase used has become progressively smaller. Small size of these particles leads to a considerable resistance to

© NCERT not to be republished

solvent flow. The mobile phase has to be pumped through the column under high pressure.

Gas Chromatography (GC): The gas chromatography (GC) is widely used for separation and purification of volatile components such as alcohols, ketones, aldehydes and many other organic and inorganic compounds. In gas chromatography, the mobile phase is gas. Gas chromatography is used widely, however, liquid chromatography is of great relevance to bioprocessing.

12.6 Spectroscopy

Spectroscopy is a technique used for the study of interaction of electromagnetic radiation with matter. It is used for the identification of substances through the spectrum emitted from or absorbed by them. It is used for the estimation of concentration of coloured compounds, analysis of chemical structure of molecules, types of molecular changes occurring during various enzymatic reactions; and in the study of intermolecular bonding. List of various spectroscopic techniques is given in Table 12.1. Spectrophotometric techniques offer a very rapid and convenient means of qualitative and quantitative estimation of biomolecules. Spectrophotometer consists of the following parts: light source, which provide the desired wavelength of light, collimator, which transmits a straight beam of light, monochromator, which splits the light into its component wavelength, and wavelength selector, that transmits only the desired wavelength (Fig. 12.11).

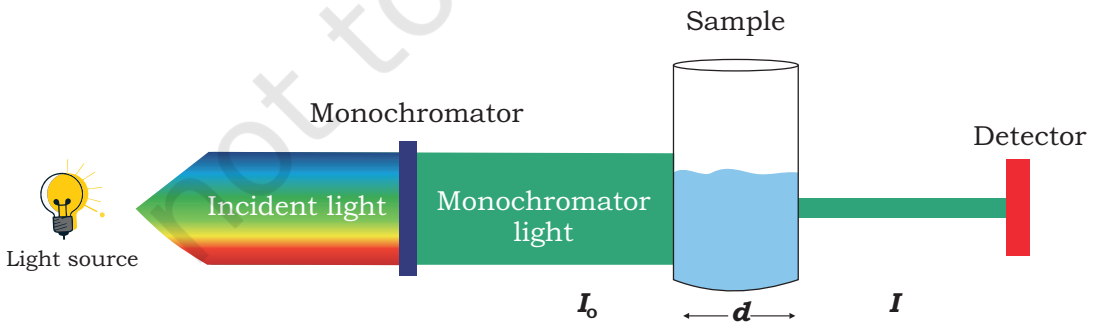


Fig. 12.11: Diagrammatic representation of instrumentation of spectroscopy

© NCERT not to be republished

Box 2

The Beer-Lambert Law

The Beer-Lambert law states that the absorbance of a solution is directly proportional to the concentration of the absorbing substance in the solution and the path length. For this reason, Beer-Lambert law can only be applied when there is a linear relationship. Beer's law is written as:

$$A = \varepsilon lc$$

Where *A* is the measure of absorbance (no units),
ε is the molar extinction coefficient or molar absorptivity (or absorption coefficient),
l is the path length and
c is the concentration.

Transmittance is the relationship between the amount of light that is transmitted to the detector once it has passed through the sample (I) and the original amount of light (I_0). This is expressed in the following formula.

$$T = I / I_0$$

Where T is the transmittance,

I is the intensity of the light coming out of the sample, and
I_0 is the intensity of the incident light beam.

Absorbance equals the negative log of transmittance and this relationship between transmittance (T) and absorbance (A) can be expressed as:

$$A = - \log (T)$$
$$A = - \log (I/I_0)$$
$$A = \log (I_0/ I)$$

Therefore,

$$A = \varepsilon lc = \log (I_0/ I)$$

If you know the absorption coefficient for a given wavelength, and the thickness of the path length for light transmitted through the solution, you can calculate concentration.

© NCERT not to be republished

Table 12.1 List of various spectroscopic techniques

Types of Energy Transfer	Spectroscopic Technique	Region of Electromagnetic Spectrum	Application
Absorption	UV/Visible spectroscopy	UV/Visible	Detection of functional groups, extent of conjugation, and determines the configurations of geometrical isomers
	Atomic absorption spectroscopy	UV/Visible	Determines the amount of various levels of metals and other electrolytes within samples
	Infrared spectroscopy	Infrared	Determines the functional groups
	Raman spectroscopy	Infrared	Contaminant identification, gemstone and mineral identification
	Nuclear magnetic resonance spectroscopy	Radio wave	Provides information about the structure and chemical environment of atoms
	X-ray absorption spectroscopy	X-ray	Determines the elemental composition and chemical bonding of molecules
Emission	Atomic emission spectroscopy	UV/Visible	Detection of trace metals, minerals, sodium, potassium and lithium
	Mass spectrometer		Analysis of proteins, peptides, checking water quality and food contamination
Photolumi-nescence	Fluorescence spectroscopy	UV/Visible	Detection of many of organic compounds, numerous aromatic active substances in drug
	Phosphorescence spectroscopy	UV/Visible	-

12.6.1 Colorimetry

Colorimetry technique utilises the interaction of light energy with coloured solutions. This instrument is used to measure transmittance and absorbance of light passing through liquid sample. It measures the intensity of the color that develops upon adding a specific reagent into a solution. The intensity of color is directly proportional to the concentration of the compound being measured. Wavelength is selected using coloured filters which absorb all but a certain limited range of wavelength. This limited range is known as **bandwidth** of the filter. The three main components of colorimeter are light source

© NCERT not to be republished

(tungsten-filament lamp), filter, cuvette containing sample and a photocell for detecting the transmitted light (light passed through the solution) (Fig. 12.12). The principle of colorimeter is based on **Beer-Lambert's law** (Box 2) which is a combination of two laws, each dealing separately with absorption of light related to the concentration of absorber and the path length or thickness of the absorbing medium. Colorimeter is inexpensive, easily transportable and used for quantitative analysis of colored compounds.

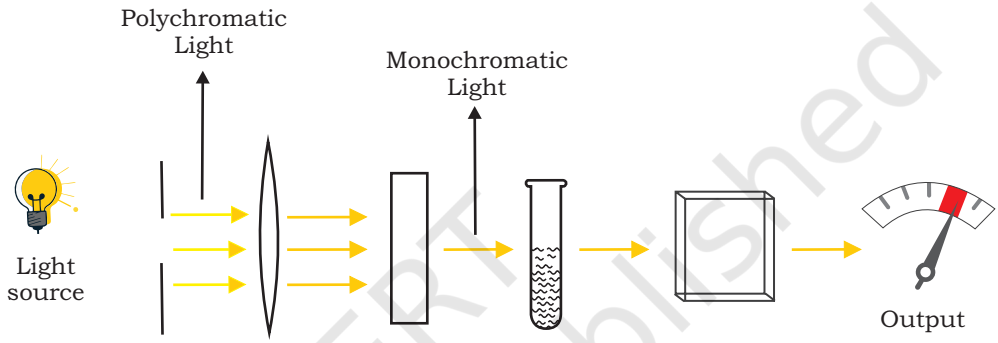


Fig. 12.12: Components of colorimeter

12.6.2 UV-visible Spectrophotometry

This instrument is used to measure the amount of light absorbed at each wavelength of UV and visible regions of electromagnetic spectrum. A spectrophotometer is a sophisticated type of a colorimeter where monochromatic light is provided by grating of a prism. In a colorimeter, filters are used which allow a broad range of wavelengths to pass through, whereas in the spectrophotometer a prism (or) grating is used to split the incident beam into different wavelengths. A 'photometer' is a device for measuring light,

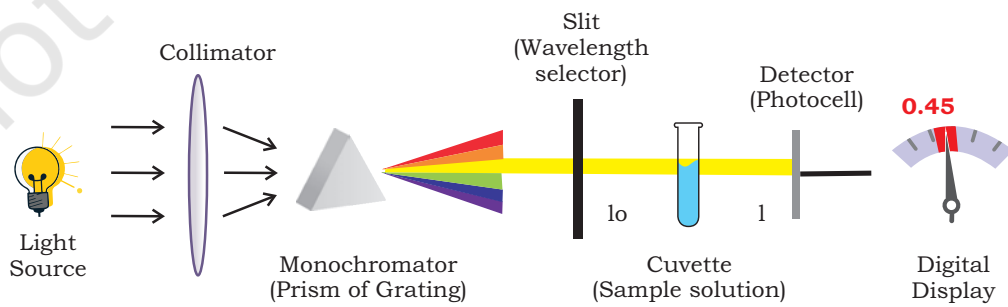


Fig. 12.13: UV-visible spectrophotometer

© NCERT not to be republished

and 'spectro' means the whole range of continuous wavelength that the light source is capable of producing. Spectrophotometer is made up of different components, such as light sources (UV and visible), wavelength selector (monochromator), sample containers (cuvette), detector, signal processor etc. (Fig. 12.13).

12.7 Mass Spectrometry

Mass spectrometry is used to identify unknown compounds, to quantify unknown materials, and to elucidate structure and chemical properties of molecules. The complete process involves the conversion of the sample into gaseous ions by electron ionisation with or without fragmentation, which are then characterised by their mass-to-charge ratios (m/z) and relative abundances. The unit of measurement of mass is Dalton (Da for short form). One Dalton is equal to 1/12th of the mass of single atom of the isotope of carbon-12. Three major components of mass spectrometry are: **ion source** for producing gaseous ions from the substance being studied, **analyser** for resolving the ions into their characteristic mass components according to their mass-to-charge ratio, **detector system** for detecting the ions, and recording the relative abundance of each of the resolved ionic species (Fig. 12.14).

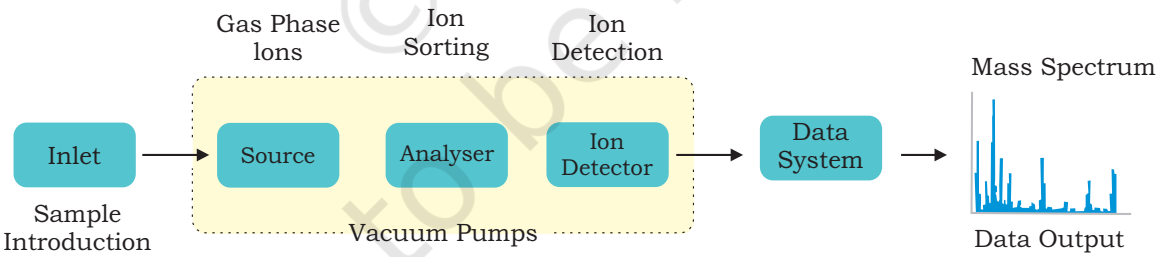


Fig. 12.14: Mass spectrometer

12.8 Fluorescence In Situ Hybridisation (FISH)

Fluorescence *in situ* hybridisation (FISH) is a cytogenetic (study of chromosomes number and structure) technique that uses fluorescent molecule that binds to highly complementary region of chromosome. FISH is useful to identify where a particular gene falls within an individual's chromosomes. Fluorescence microscopy is

© NCERT not to be republished

used to find the location of the fluorescent molecule in the chromosome. It is an important tool to understand chromosomal abnormalities.

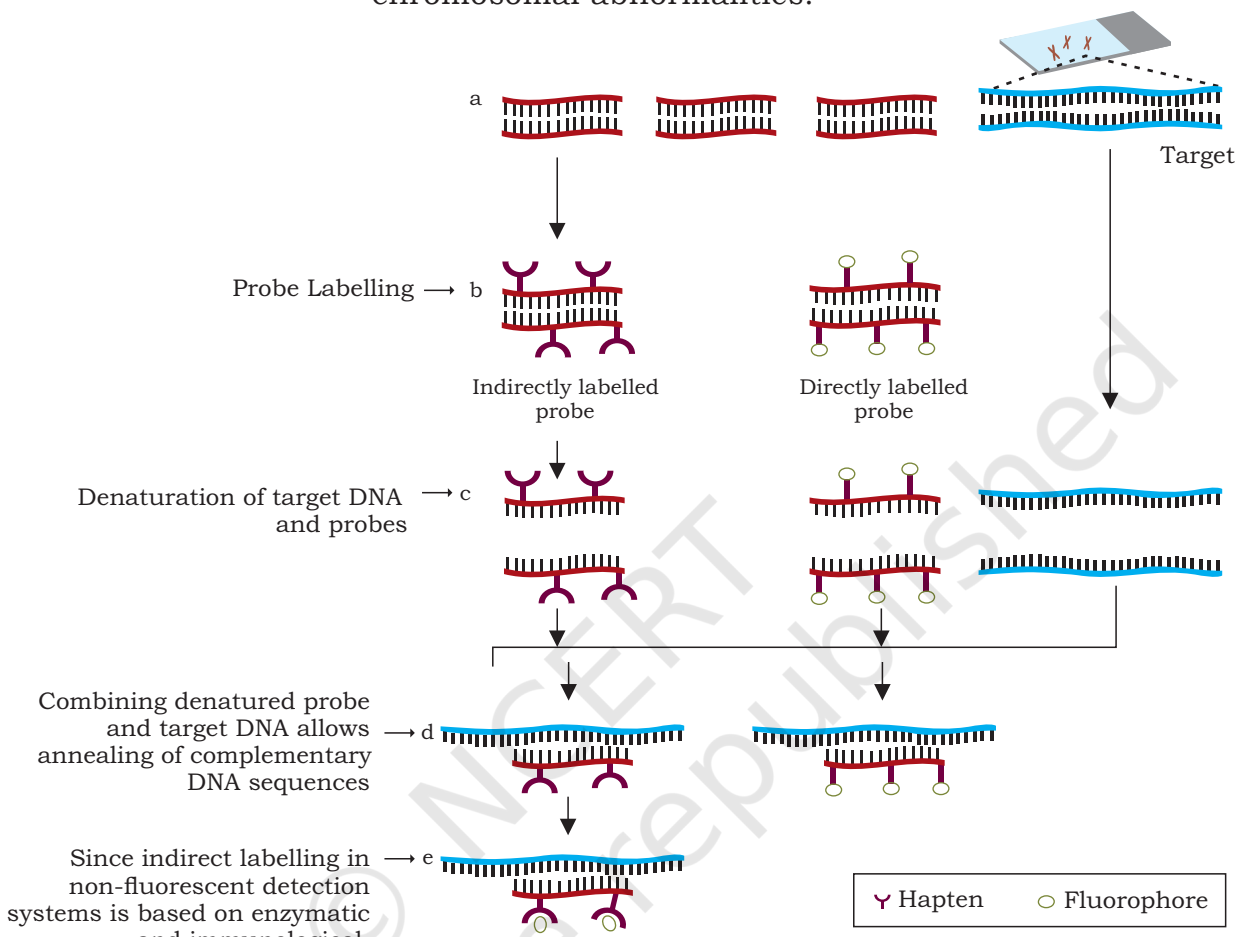


Fig. 12.15: Basic steps of FISH

The main components of FISH are:

(i) A fluorescent DNA molecule (probe), and

(ii) The chromosome (target sequence)

Here is how FISH works

1. Design a molecule (probe) complementary to the known sequence. The probe is labelled with a fluorescent molecule e.g. fluorescein (Fig. 12.15).
2. Put the chromosomes on a microscope slide and denature them.
3. Denature the probe and add it to the microscope slide.
4. The probe hybridises to its complementary site.

© NCERT not to be republished

5. The excess probe is washed off and the chromosome is observed under a fluorescent microscope. The probe will show as one or more fluorescent signals in the microscope, depending on how many sites it can hybridize to.

Application of FISH

There are various applications of FISH including chromosome painting. For chromosome painting, multifluor FISH probe can be used to generate a karyotype in which each chromosome appears to be painted with a different colour (Fig. 12.16).

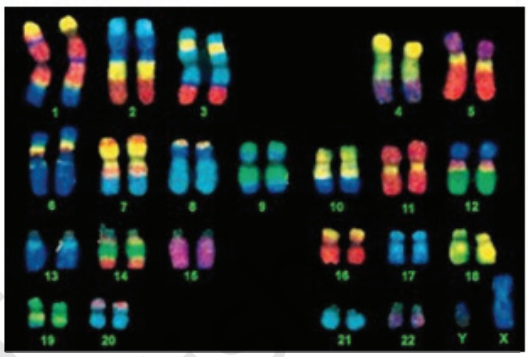


Fig. 12.16: Chromosome painting

First, a collection of DNA sequences are prepared for using as probe for each chromosome. Then these DNA sequences are labelled with combinations of fluorochromes that produce a unique color. The fluorescent DNA probes and metaphase chromosomes are mixed together; then the hybrids are visualised under fluorescent microscope.

12.9 DNA Sequencing

As you are aware, DNA is made up of four types of nucleotides (A,T,G,C), which are linked together through phosphodiester linkages, and is the carrier of genetic information. DNA sequencing refers to finding the order of nucleotides (ATGC) on a piece of DNA. The order of these four bases is key to the unique feature of the functional unit of a given DNA (e.g., a gene). In other words, the sequence of DNA comprises the heritable genetic information that forms the basis for the developmental programs of all living organisms. The advent of DNA sequencing has significantly accelerated biological research and discovery.

The unique order of these bases greatly influences health, e.g., what disease a person is prone to and how the person will react to different medications. Understanding a particular DNA sequence can shed light on a genetic condition (e.g., disease) and offer hope for the eventual development of treatment. Thus, an alteration in a DNA sequence can lead to an altered or non functional protein, and hence to a harmful effect. Also, in order to understand the structure, function and evolutionary history of a cloned

© NCERT not to be republished

DNA, its primary structure, i.e., the nucleotide sequence is required. DNA sequencing technology is also extended to environmental, agricultural and forensic applications. Thus, determining the DNA sequence is useful in basic research studying fundamental biological processes, as well as in applied fields such as diagnostic or forensic research. The rapid speed of sequencing attained with modern DNA sequencing technology has been helpful in the sequencing of the complete DNA sequences of many animal, plant, and microbial genomes including that of the human.

12.9.1 DNA sequencing methods

Historically there are two main methods of DNA sequencing, namely, (1) Enzymatic method (Sanger's method, dideoxynucleotides chain termination method) and (2) Chemical degradation method (Maxam and Gilbert method).

The two methods are described in detail in the following sections:

(a) Sanger's method — Sanger's method works on the principle that dideoxynucleotides (dideoxyadenine,

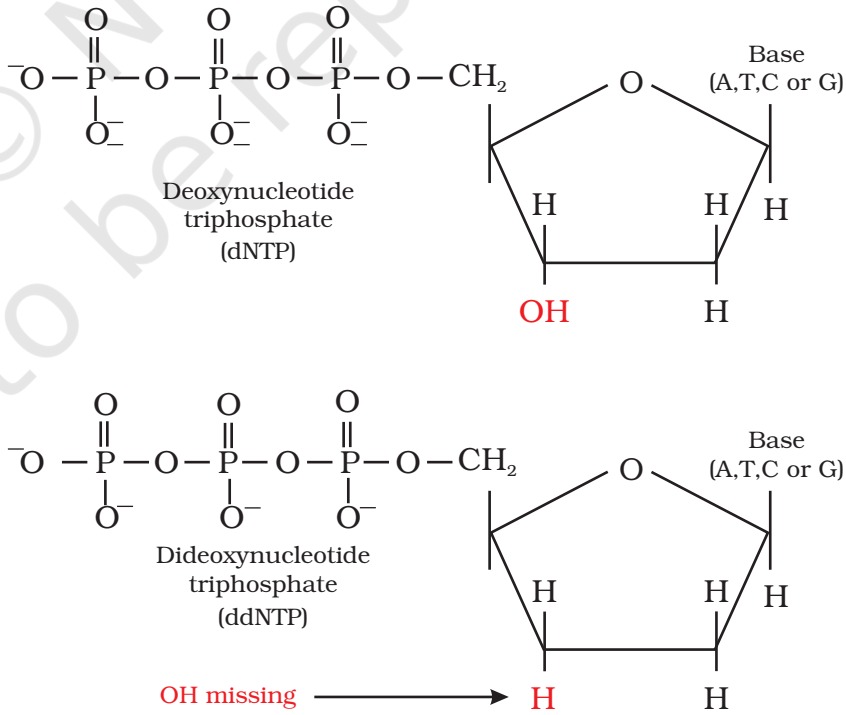


Fig. 12.17: Structure of normal deoxynucleotide (top) dideoxynucleotide (bottom)

© NCERT not to be republished

dideoxyguanine, etc., which resemble normal nucleotides but lack the normal-OH group at 3' position) get incorporated, instead of the normal deoxynucleotide, into the newly synthesised chain (daughter chain) which leads to termination of synthesis of the new strand at that point (and hence it is called as chain termination method) (Fig. 12.17). Sanger's method is considered as a gold standard for DNA sequencing. It is used even today for routine sequencing applications and also for validation of Next Generation Sequencing (read in following sections of this chapter) data.

In Sanger's method of DNA sequencing, the template DNA (the sequence of which has to be determined) is mixed with a primer (a small piece of chemically synthesised DNA

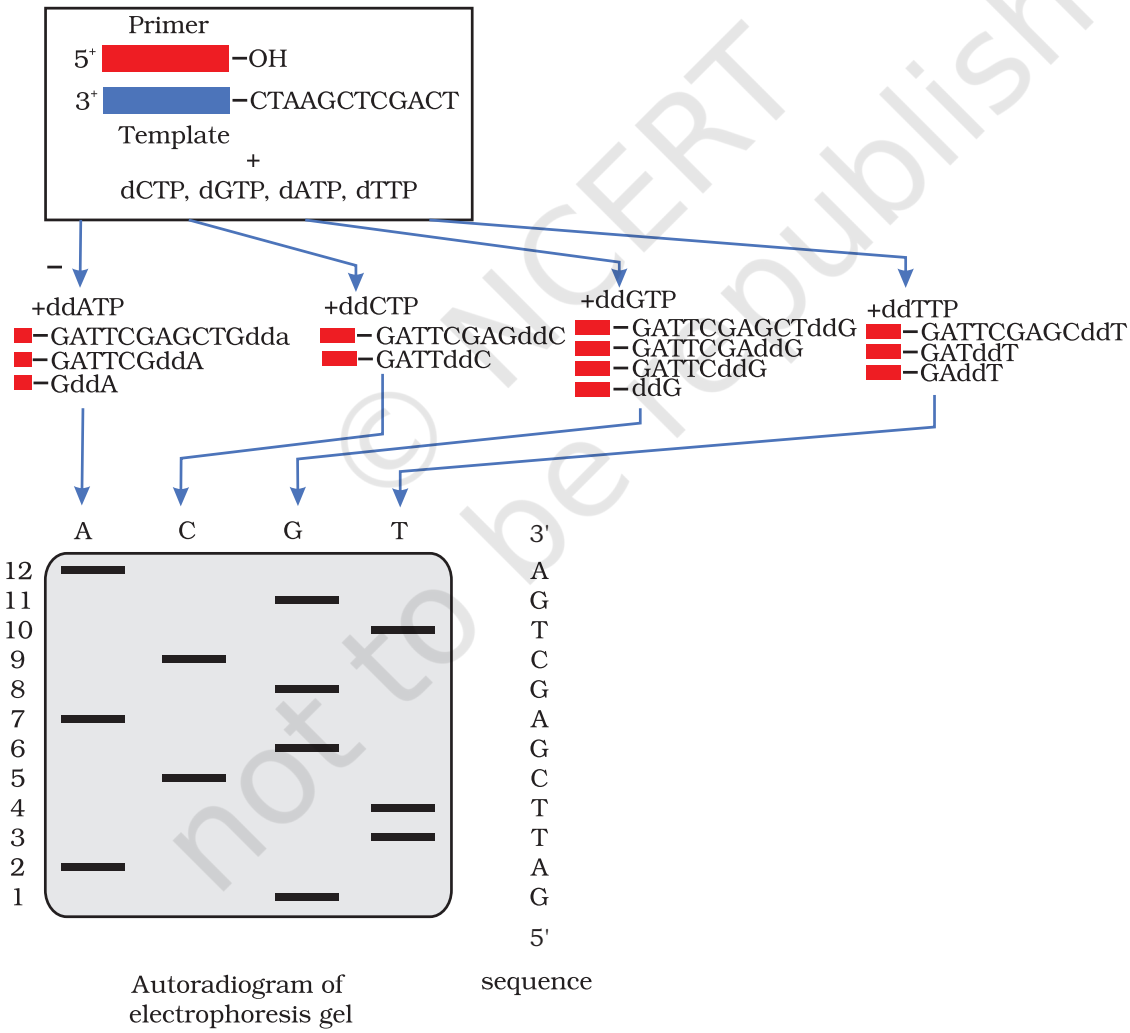


Fig. 12.18: Sanger's method of DNA sequencing

© NCERT not to be republished

of defined sequence that can pair with the template DNA to act as a starting point for DNA synthesis) complementary to the template DNA and the four normal dNTPs, one of which is labelled radioactively or fluorescently. This mixture is then split into four different tubes that are labelled A, C, G, and T. Each tube is then 'spiked' with a different ddNTP (ddATP for tube A, ddCTP for tube C, ddGTP for tube G, or ddTTP for tube T). Following this, DNA polymerase is added, and using the DNA template and its complementary primer, the synthesis of new strands of DNA complementary to the template begins. Occasionally, a dideoxynucleotide is added instead of the normal deoxynucleotide and synthesis of that strand is terminated at that point. Thus, all fragments in lane A will end in an A, fragments in lane C will all end in a C, fragments in lane G will all end in a G, and fragments in lane T will all end in a T. After carrying out the reaction for a fixed time the newly synthesised DNA strands (fragments) are separated through high resolution polyacrylamide gel electrophoresis (PAGE) and the DNA fragments are visualised by exposing the gel to X-ray film for nucleotides labelled radioactively. Finally, the sequence of the DNA is read from the gel by starting from the bottom and reading upward on X-ray film (Fig. 12.18).

(b) Chemical Degradation Method (Maxam and Gilbert Method)— In the chemical degradation method of DNA sequencing, the DNA fragment, whose sequence is to be determined is cleaved in a base specific manner using chemicals as shown in Fig.12.19. Before the chemical mediated degradation of the DNA, the DNA is labelled at the 5' end using enzyme polynucleotidyl

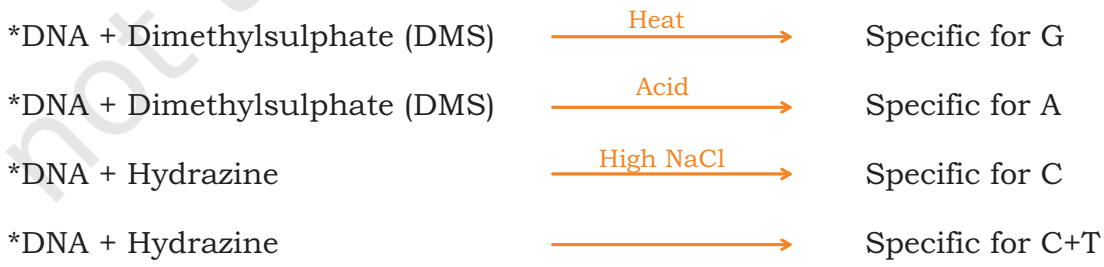


Fig. 12.19: Base specific cleavage reactions used in Maxam and Gilbert method

© NCERT not to be republished

kinase and gamma-^{32}P labelled nucleotide. The fragments generated are subsequently separated using high resolution PAGE to resolve the sequence order. These gels are placed under X-ray film, which then yields a series of dark bands which show the location of radiolabeled DNA fragments. The fragments are ordered by size and therefore, we can deduce the sequence of the DNA molecule.

12.9.2 Automated DNA sequencing

Majority of DNA sequencing is now done through automated method. These automated sequencers primarily work on Sanger method of DNA sequencing. In automated sequencing, fluorescent labelled dideoxynucleotides are used which has eliminated the need for radioactive isotopes. The slab gel has been replaced by polymer filled capillary tubes in automated equipment. Automation of DNA sequencing has made the method much quicker and more reliable. For example, in one year, by using manual sequencing, on an average, a person can sequence 20,000 to 50,000 bases while automated sequencer can sequence that long in just a few hours. Furthermore, total cost of material for one gel using the automated method is approximately half, compared to that of the manual method. In automated DNA sequencing, all four dideoxy reactions are carried out in a single tube, which is possible because each dideoxynucleotide is labelled with a different fluorescent dye such as Rhodamine 110 (RHO)-A (which gives green fluorescence), Rhodamine 6G (REG)-C (which gives blue fluorescence), Tetramethyl Rhodamine (TAMRA)-G (which gives black fluorescence) and X-Rhodamine (ROX)-T (which gives red fluorescence). The contents of the single tube reaction are loaded onto a single lane of a gel and electrophoresis is done. The sequence is determined by the order of the dyes coming off the gel. A fluorimeter and computer are hooked up to the gel and they detect and record the fluorescent dye attached to the fragments as they come off the gel. All the DNA fragments labelled with fluorescent dideoxynucleotide are 'read' by a laser and the fluorescence intensity translated into a data 'peak' (Fig. 12.20).

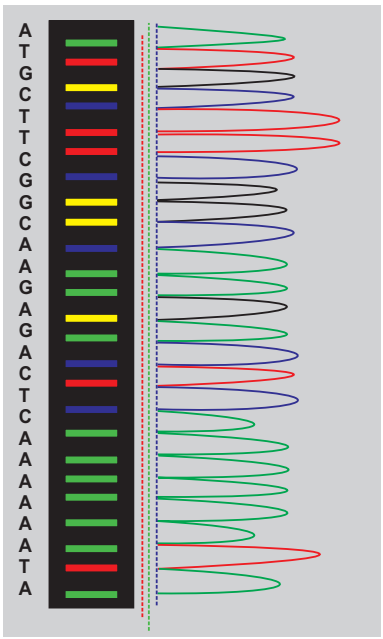


Fig. 12.20: Chromatogram of automated sequencing

© NCERT not to be republished

Box 3

Next-generation Sequencing (NGS)

NGS methods have facilitated the sequencing of very large DNA such as whole genome as the above mentioned methods work optimally for short sequence analysis up to few Kb size DNA and it is difficult to analyse the whole genome as it will take too much time and cost. The evolution of DNA sequencing from the historical methods of Sanger (Sanger sequencing) and Maxam & Gilbert (Maxam-Gilbert sequencing) to today's high-throughput technologies has occurred at a breathtaking speed. In the past 30 years or so, these high-throughput technologies have given rise to super-exponential growth in sequence data generation and the resultant data have led to transformative applications ranging from basic biology to criminal investigation and prenatal diagnostics.

NGS allows massively parallel sequencing reactions and therefore, they are capable of analysing millions or even billions of sequencing reactions at the same time. The widely used NGS platforms are Roche/454 sequencing, Solexa/Illumina and SOLiD platforms.

12.10 DNA MICROARRAY

DNA microarray technology is a high throughput hybridisation-based technique that is used to analyse a large number of DNA fragments in parallel for quantification of the expression of large number of genes. It uses the property of two DNA strands to pair with each other by

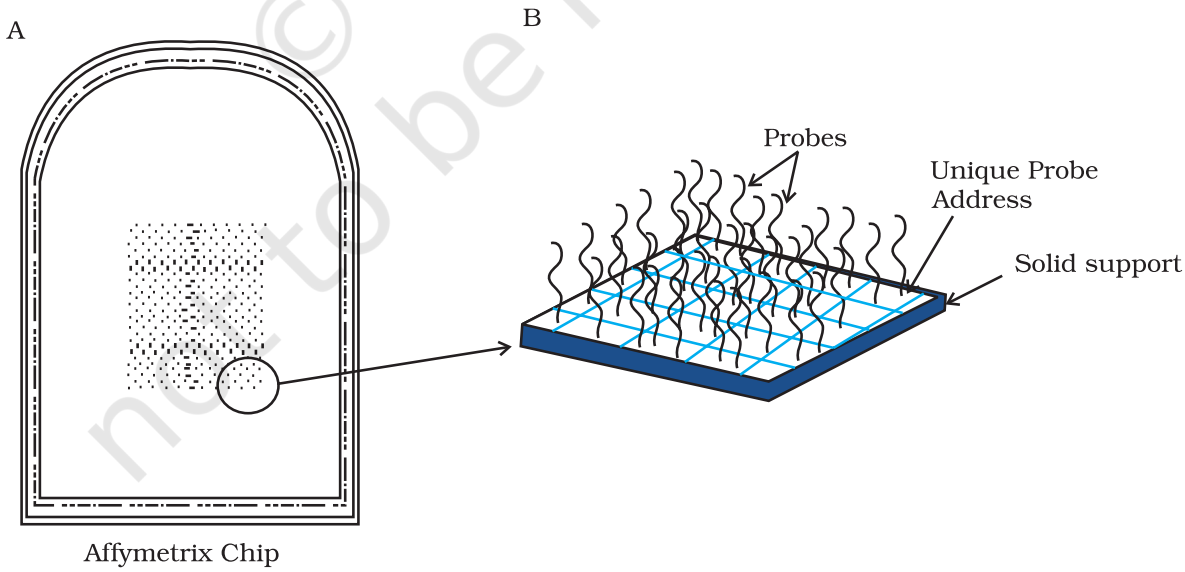


Fig. 12.21: DNA Microarray chip

© NCERT not to be republished

forming hydrogen bonds between complementary nucleotide bases. Hence, the principle of DNA microarray technology is that complementary DNA sequences can be used to hybridise to immobilised DNA fragments on the chip and individual hybridisation events can be recorded.

Thousands of single stranded DNA (ssDNA) segments corresponding to the gene transcripts (mRNA) or other genomic regions, are immobilised on a small solid surface and are referred as **microarray chips** (Fig. 12.21). These chips are usually made of either glass or nylon, and are coated with a special surface coating that allows spotting of DNA on the chip. ssDNA segments immobilised on the chips are called as **probes** and are arranged in rows and columns on the chip. This arrangement of the probes helps to find the location of any specific fragment on the

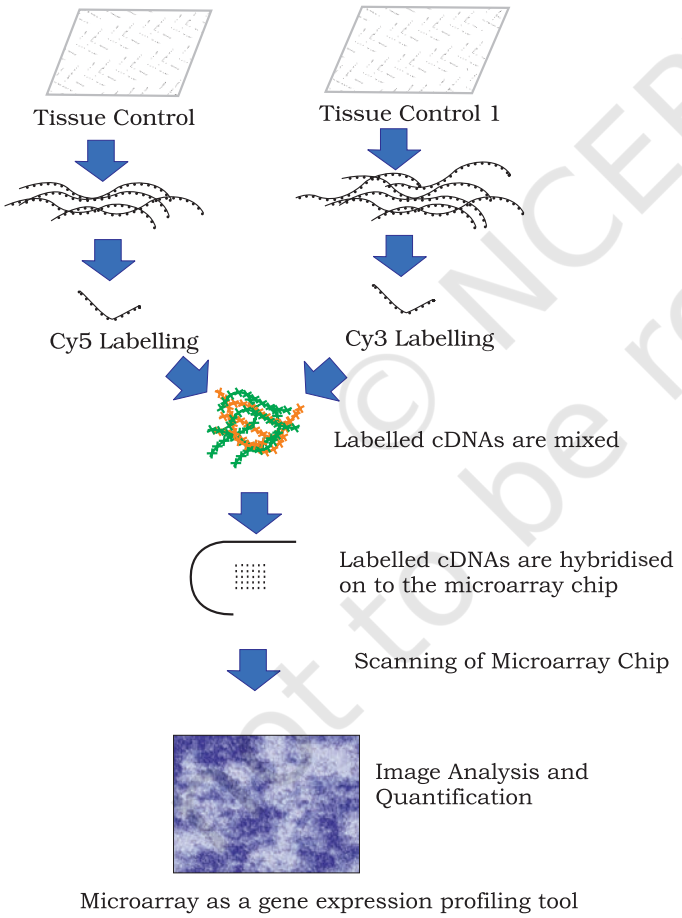


Fig. 12.22: An example of a cDNA microarray experiment

© NCERT not to be republished

chip. Usually probes are either cDNAs, PCR amplicons or oligonucleotides that correspond to the mRNAs and are referred as cDNA or oligonucleotide probes. Oligonucleotide probe-based arrays are very popular. These probes are short sequences that are complementary to the known/predicted transcripts from a single species, and allows to analyse the expression of thousands of genes in parallel.

A typical microarray experiment includes the following steps (Fig. 12.22).

1. Extraction of mRNA
2. Probe labelling
3. Hybridisation and washing
4. Scanning and data analysis

Extraction of mRNA

In a cell, transcription from a gene produces a mRNA, which makes proteins by translation. Depending upon the requirement, many RNA copies of the same gene are formed. Therefore, activity of a particular gene can be quantified by quantification of mRNA. Because mRNA is degraded very easily, it is converted into complimentary DNA (cDNA) through reverse transcription, which is more stable and represents actively transcribing genes in the cell.

Probe Labelling

cDNA fragments are digested with restriction endonucleases and resulting fragments are attached with fluorochrome dyes. Cy3 and Cy5 are most commonly used fluorescent dyes to this purpose. Therefore, probes are made of thousands of labelled nucleic acid fragments.

Hybridisation and Washing

Labelled DNA fragments are hybridised with the microarray chip. For hybridisation, DNA chips (having thousands of single stranded probes) are exposed and allowed to react with single stranded fluorochrome labelled DNA fragments. These fragments bind to their complementary single stranded probes on the chip and form duplexes. The number of duplexes formed is the reflection of the number of its DNA segments complementary to its probe. Other DNA fragments, which do not find their complementary

© NCERT not to be republished

probe on the chip are washed away.

Scanning and Data Analysis

Finally, microarray chip with hybridised labelled DNA fragments are scanned using a highly sophisticated scanner. Image analysis is performed using sophisticated software program that helps to determine how much labelled cDNA is bound to target probes on the chip. Unique addresses of these target probes and their association with specific genes is used to interpret and quantify data. Microarray data analysis software uses different colours to represent the expression level of genes in one condition with respect to other condition. Genes whose expression increases in one condition with respect to an other condition are called as **upregulated genes**, and other whose expression decreases are called as **downregulated genes**. Classically, green colour is used to represent upregulated genes and red colour is used to represent downregulated genes.

12.11 Flow Cytometry

There are diverse varieties of cells in organisms, which perform one or the other function. Understanding features of a specific cell, whether physical or chemical may provide valuable information from different perspectives. Such an understanding of cells based on either of the parameters mentioned earlier may be used for qualitative or quantitative measurement of cells. History of quantifying cells based on physical or chemical properties can be traced back to the Coulter counter used during late 1950s , which was invented to quantify particle in suspension based on the principle of change in impedance proportional to the particle volume. Such impedance can be detected under electric field when the particle passes through an area separated by two chambers having electrolyte solution (Fig. 12.23).

Present day's flow cytometry is also based on the same principle of impedance due to particle passing along with flowing channel of fluid. In order to achieve this, cells (sample) are passed through the flow cytometer in such a way that cells flow one by one in the fluid stream.

© NCERT not to be republished

Laser beams fixed in the passage detects cells one by one based on its properties or the dye used to label the same. The deflected light is then detected by a sensor, which based on the intensity of light source and the deflected light provides information about particle or cell. These detectors may either be in the line of the light beam to detect the surface property or volume, or it may be perpendicular, which can detect the internal properties. Signals are received by sensor and ultimately an image of the object emerges comprehensively (Fig. 12.23). Similarly, cells present in the immune system may have thousands of antigens on their surface. In order to locate any antigen or protein in a cell, specific antibodies are used quite often. Fluorescence dye is commonly used to label antibodies for the purpose of easy identification and localisation. Different cells of a

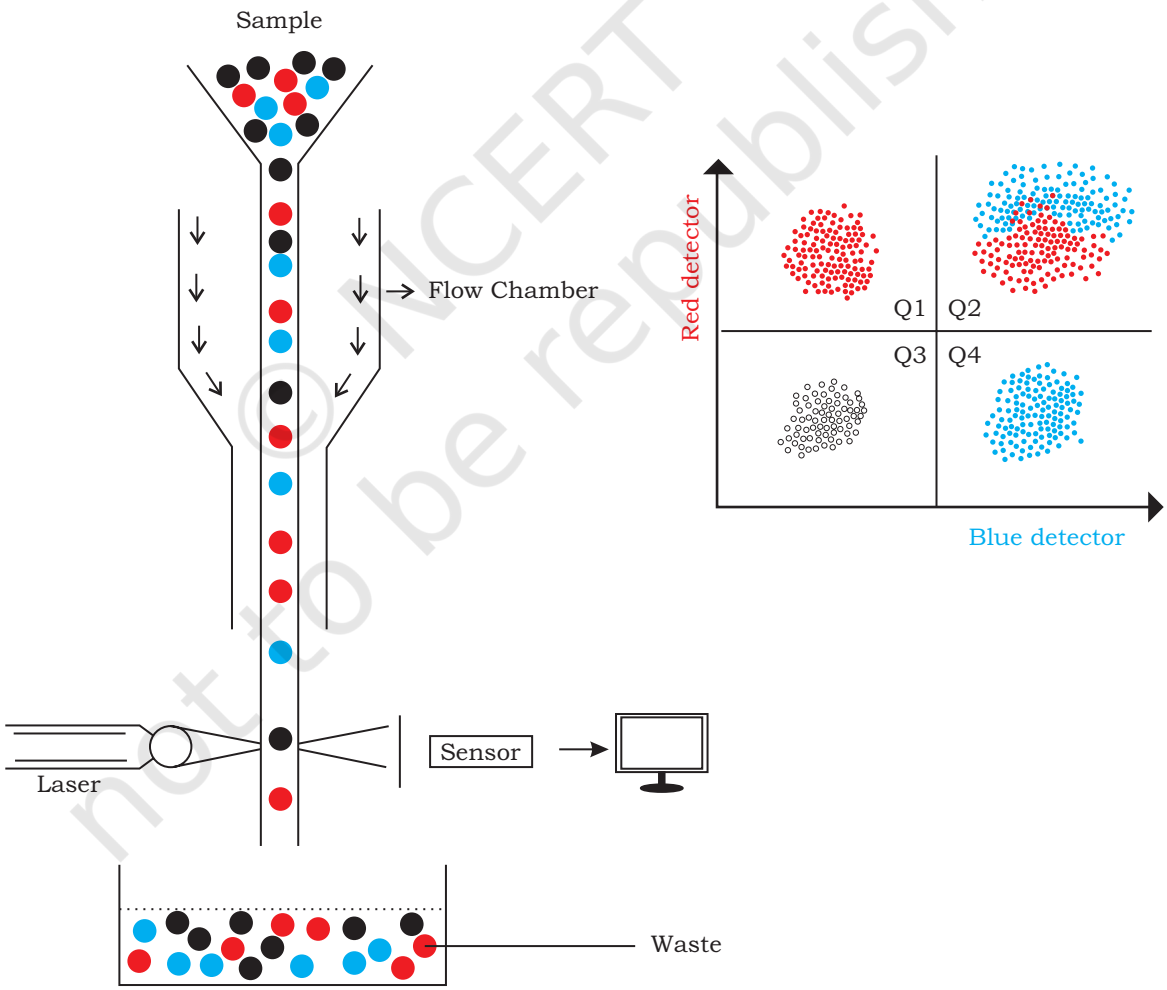


Fig. 12.23: A diagrammatic representation of flow cytometry

© NCERT not to be republished

mixture activated with different fluorescent dye can also be mechanically separated by this technique called **Flow Cytometry**.

Sometimes, cells labeled with different fluorescent dyes in a mixture is detected on the basis of fluorescence present on it. For the purpose of separation of cells, first the mixture of fluoresce labeled cells are loaded on a charging electrode followed by its release drop-wise. Laser based sensor present in the path of dropped cells detect these based on their fluorescent label and the information is recorded in a computer, which can be seen as the plot shown in Fig.12.23.

The development of modern, rapid and sophisticated biological tools and techniques has made biological studies more accurate, fast, quantitative and reproducible.

SUMMARY

- Simple light microscopy enables us to see things that are otherwise too small to be observed by our naked eyes. Studying minute details of internal organisation of cells is so very diverse that it requires more refined microscopy like contrast microscopy or electron microscopy.
- Different types of centrifuges like differential centrifuge, high speed centrifuge, density gradient centrifuge and ultracentrifuge helps in the separation of various biomolecules present within cells based on their densities under the influence of gravitational force and spinning them in a solution around an axis at high speed using centrifugal force.
- By means of electrophoresis, many important biological molecules like DNA, RNA and proteins can be separated and studied on the basis of charge to mass ratio of macromolecules under the influence of an electric field.
- Enzyme-linked immunosorbent assay (ELISA) is a highly sensitive and quantitative immunological assay for measurement of antigen and antibody concentration in given sample. Different types of ELISA like direct, indirect, sandwich and competitive are used in diagnosis and scientific studies.
- Various different chromatography methods: Adsorption chromatography, Ion-exchange chromatography, affinity chromatography and gas chromatography are exploited for purification of proteins, peptides, amino acids, sugars, nucleic acids, alkaloids, vitamins and steroids.

© NCERT not to be republished

- Similarly, to elucidate the chemical structure of molecules, spectroscopy techniques are used. Colorimetry technique measures the transmittance and absorption of light passing through liquid sample and measures the concentration of the sample.
- Fluorescence In Situ Hybridisation (FISH) technique uses fluorescent molecules binding to highly complementary regions of chromosome and facilitates in the identification of a particular gene in an individual chromosome and hence, play crucial role to understand chromosomal abnormalities.
- Sequence of DNA comprises the heritable genetic information that forms the basis for the developmental programs of all living organisms. The advent of DNA sequencing has significantly accelerated biological research and discovery. Sanger method of DNA sequencing are developed about four decades earlier is even used today for routine sequencing applications. Many other sequencing methods called Next Generation Sequencing are available these days such as; Roche/454, Solexa/Illumina and SOLiD plateforms.
- DNA microarray analysis assist in analysing expression levels of large number of genes.
- In flow cytometry, cells pass through a laser beam, allowing their physical and chemical properties to be analysed.

EXERCISES

1. The function of ethidium bromide in electrophoresis is to
   (a) track the progression of electrophoresis
   (b) visualise the DNA molecules
   (c) separate the DNA molecules
   (d) provide charge to DNA molecules
2. Match the following

	Column I	**Column II**
(a)	Separation of ionic solutes	Affinity chromatography (AFC)
(b)	Separation of biomolecules with different binding specificities	Gas chromatography (GC)
(c)	Separation of volatile components	Ion-exchange chromatography (IEC)

© NCERT not to be republished

3. Mass spectrometry is used to
   (a) identify unknown compounds
   (b) elucidate the structure of molecules
   (c) quantify compounds
   (d) All of the above

4. Match the following table with reference to Antigen

	ANTIGEN	ANTIBODY	PROCEDURE
(i)	Free	Bound to surface	Direct ELISA
(ii)	Bound	Only one labeled primary antibody used	Indirect ELISA
(iii)	Bound	Labeled secondary antibody used	Sandwich ELISA

5. In DNA gel electrophoresis,
   I. Longer DNA fragments remain close to the well.
   II. Longer DNA fragments move towards the positive end of gel.
   III. Smaller DNA fragments move close to the positive end of gel.
   IV. Smaller DNA fragments remain close to the well.
   Which of the above options are correct
   (a) I and III
   (b) II and IV
   (c) Only II
   (d) None of the above

6. For a resolved image of the surface of an object, which of the following microscopes would you prefer
   (a) Transmission electron microscope
   (b) Scanning electron microscope
   (c) Phase contrast microscope
   (d) Fluorescence microscope

7. Match the following:

	Column I	Column II
(a)	Engvall and Perlman	Microscopy
(c)	Robert Hooke	DNA sequencing
(c)	Sanger	ELISA

8. Which of the following techniques is feasible to quantify the expression of a large number of genes
   (a) Mass spectrometry
   (b) Microarray
   (c) FISH
   (d) Agarose gel electrophoresis

© NCERT not to be republished

9. Differentiate between the following types of microscopy techniques
   (a) Scanning electron microscopy (SEM) and transmission electron microscopy (TEM)
   (b) Dark field microscopy and bright field microscopy
   (c) Phase contrast microscopy and confocal microscopy
10. Discuss the principle of agarose gel electrophoresis.
11. Name a tracking dye which is used to track DNA as well as proteins during electrophoresis. What will happen if you forget to add tracking dye to your sample during electrophoresis?
12. Two polyacrylamide gels A and B were prepared. Gel A had 4% acrylamide whereas Gel B had 12% acrylamide. Based on the given information answer the following
   (a) Which gel is harder: A or B?
   (b) Which gel offers greater friction to the proteins: A or B?
   (c) Which gel (A or B) will be used to separate a mixture containing low molecular weight proteins?
   (d) Which gel (A or B) will be used to separate a mixture containing both low and high molecular weight proteins?
13. What is a chromatogram? Draw a well labeled diagram of a chromatogram of a mixture containing three different solutes.
14. Explain the principle of FISH. How is FISH technique applied in chromosome painting? What are the advantages of chromosome painting?
15. Mention the various applications of spectroscopy techniques.
16. What are major components of UV-visible spectrophotometer? Explain each in brief.
17. Write the major differences between the Sanger's method and Maxam and Gilbert's method of DNA sequencing.
18. Write the principle of flow cytometry.

© NCERT not to be republished